APRIL
1955
VOLUME 19
NUMBER 2

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

Published by the American Psychological Association, Inc.
at Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth St. N.W., Washington, D.C.

Edited by Laurance F. Shaffer

Teachers College, Columbia University

Advisory Editors: Edgar A. Doll, *Bellingham, Wash.* Ann Magaret Garner, *University of Illinois.* William A. Hunt, *Northwestern University.* E. Lowell Kelly, *University of Michigan.* Morris Krugman, *Board of Education, New York, N. Y.* Bertha M. Luckey, *Board of Education, Cleveland, Ohio.* Fred McKinney, *University of Missouri.* Anne Roe, *New York, N. Y.* Carl R. Rogers, *University of Chicago.* Nevitt Sanford, *University of California.* Edward Joseph Shoben, Jr., *Teachers College, Columbia University.* J. R. Wittenborn, *Rutgers University.*

Lorraine Bouthilet, *Managing Editor*

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

Editorial Office

Manuscripts submitted for publication and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y. All manuscripts should be submitted *in duplicate* to facilitate referral to the Advisory Editors. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

Style of manuscripts. Manuscripts should be prepared in the style described in the Publication Manual of the American Psychological Association (*Psychological Bulletin*, 1952, 49, 389–449). The Manual may be purchased from the Washington office of the Association for $1.00. Otherwise acceptable manuscripts which depart from the standards of the Manual are returned for revision and retyping.

Brief Reports. A one-page condensation of an acceptable research report may be given early publication without charge if the author agrees (*a*) to make 100 mimeographed copies of the full report available to those who request it, (*b*) to send a copy of the full report to the Editor for deposit in the American Documentation Institute, and (*c*) not to have the full report published elsewhere. For detailed instructions on how to prepare a Brief Report, see the back of the *Contents* page of this Journal or write to the Editor.

Early Publication. Authors may secure the early publication of an acceptable article in the first issue to go to press after its acceptance by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. The charge to the author for such early publication is $20.00 per printed page.

Reprints. Authors receive fifty reprints gratis of all articles except early publication articles and Brief Reports. Additional quantities may be ordered when returning proof.

Business Office

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must reach the subscription office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $5.00, foreign $7.50, single number $1.50.

Entered as second-class matter March 21, 1955, at the post office at Lancaster, Pa., under the act of March 3, 1879. Acceptance for mailing at special rate of postage provided for in Section 538, Act of February 25, 1925, authorized May 27, 1940.

Copyright, 1955, by the American Psychological Association, Inc.

JOURNAL OF CONSULTING PSYCHOLOGY

April, 1955 Vol. 19, No. 2

Contents

Correspondence Between Attitudes and Images of Parental Figures in TAT Stories and in Therapeutic Interviews: *Mortimer M. Meyer and Ruth S. Tolman* - - - - - - - 79

The Application of a Psychometric Scoring Procedure to a Group Modification of the Thematic Apperception Test (N-TAT): *Leonard D. Eron, Florence Sultan, and Frank Auld, Jr.* - - - - - - - - - - - - 83

The Effects of Sexual Stimulation Under Varying Conditions on TAT Sexual Responsiveness: *Paul H. Mussen and Alvin Scodel* - - - - - - - - - - - - - - - - - - 90

The Diagnostic Significance of Usual and Unusual TAT Stories: *Jerome S. Miller and Alvin Scodel* - 91

The Diagnostic Discrimination of the "Basic Rorschach Score": *Ben C. Finney* - - - - - 96

Psychotherapeutic Technique, Professional Affiliation, and Experience Level: *Hans H. Strupp* - 97

Adjustment Changes in the Absence of Psychotherapy: *Rosalind F. Dymond* - - - - - 103

Language and the Interpretation of "Sexual" Symbolism: *Harold H. Mosak* - - - - - - 108

Conformance in Role Perception as a Personality Variable: *Theodore R. Sarbin and Curtis D. Hardyck* - - - - - - - - - - - - - - - - - 109

The Rorschach Experience Balance in Epileptics: *Paul W. Pruyser and Angela T. Folsom* - 112

A Note on Some Projected Familial Attitudes Associated with Rorschach Movement Responses: *Jerome L. Singer and Daniel A. Sugarman* - - - - - - - - - - - - - 117

Relationship Between Diagnostic Category and Deviant Verbalizations on the Rorschach: *William T. Powers and Roy M. Hamlin* - - - - - - - - - - - - - - - - - 120

The Measurement of Anxiety in Clinical Situations: *Arnold H. Buss, Morton Wiener, Ann Durkee, and Marc Baer* - - - - - - - - - - 125

Cross Validation of the Organic Brain Pathology Scale for a Test of Symbol Arrangement: *Theodore C. Kahn* - 130

Digit Symbol Performance as a Function of Increasing Levels of Anxiety: *Joseph D. Matarazzo and Jeanne S. Phillips* - - - - - - - - - - - - - - - - 131

The Discriminative Power of the Strauss-Lehtinen Battery in Adult Mental Patients: *A. A. Rafi* - 135

Normative Data on Progressive Matrices (1947): *Meredith W. Green and Josephine C. Ewert* 139

Relationships Between Verbal and Nonverbal Parts of the CTMM and WISC: *Grace T. Altus* - 143

Food Aversions: Some Additional Personality Correlates: *Wendell I. Smith, Elizabeth K. Powell, and Sherman Ross* - - - - - - - - - - - 145

The Effect of Pinhole Observation on the Flicker Fusion Frequency in Patients with Brain Pathology: *James C. Stauffacher, John B. Marks, and Albert F. Ax* - - - - - - - 150

Emotional Adjustment and Dieting Performance: *John Summerskill and C. Douglas Darling* 151

New Books and Tests - 154

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 70 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style (1). Headings, tables, and references are avoided or, if essential, must be counted in the 70 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 70-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. —— from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $—— for microfilm or $—— for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Extended report. The full report is prepared in the style specified by the *Publication Manual* (1), except that it may be typed with single spacing for economy in photoduplication by the ADI.

Reference

1. American Psychological Association. Council of Editors. Publication manual of the American Psychological Association. *Psychol. Bull.*, 1952, 49, 389–449.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Correspondence Between Attitudes and Images of Parental Figures in TAT Stories and in Therapeutic Interviews[1]

Mortimer M. Meyer and Ruth S. Tolman
VA Regional Office Mental Hygiene Clinic, Los Angeles

The assumption has been made that in the TAT, by the construction of stories around ambiguous pictures, the individual organizes material from his own experience. The expression of consciously recognized autobiographical material in TAT stories, or, as Murray (8) puts it, the narration of "frank and unabashed autobiographies," has been the subject of several investigations. One of the earliest was Murray's (9) comparison of the results of the TAT with information obtained through other devices, among them interviews and autobiographies. His results were never subjected to statistical treatment, but clearly Murray regarded them as satisfactory validation. Harrison (5) used the method of matching reports based on the TAT with case history material of the same individuals and met with more than chance success. In another experiment Harrison found that biographical information could be deduced from the stories of mentally ill patients with a fairly high degree of validity. Rotter (10) on the basis of his investigation indicates his finding that "the identified figure" (the hero) has a history similar to the subject's own.

Markmann's (6) study of the relationship between stories told by 40 individuals in response to three of the TAT pictures and their past history revealed a more faithful reflection of this history in some pictures than in others. She found that stories to Picture 1 most frequently reflected the actual history, especially if the theme of the story was an aggressive one. Tomkins (12), in reviewing Markmann's findings, suggests that much work needs to be done to explore the conditions under which past history is reflected in the TAT.

Combs (4) in a study of TAT stories and personal experience as recorded in autobiographies found that approximately 30 per cent of TAT stories contained significant material drawn from the subject's own life experience. In terms of his criterion, Picture 2 encouraged the most transfer of personal experience.

Now in all these studies, and in others reported by Bell (1, pp. 207–252), there is material on the subject's attitudes toward or experiences with the parents or with parental figures, but little or nothing in regard to the attitudes which the subject attributed to these figures. In this same way, Bellak's (2) outline for analysis or interpretation of the TAT includes a section on "Attitudes to Superior (Parental) Figures or to Society"; but again these refer to the subject's attitudes, not to those ascribed to parental figures. An unpublished Ph.D. dissertation by Butler (3), "Parent Figures in TAT Records of Children in Disparate Family Situations," may deal with some aspects of this question for children, but does not touch on the situation for adults.

There are two recent studies, Mitchell's (7) and Singer's (11), which do involve attitudes ascribed to parents. Both use the TAT and both assume that the TAT stories correctly represent the subject's feelings about his parents' attitudes toward him. These studies are concerned primarily with familial attitudes

[1] Presented at the meeting of Western Psychological Association, Long Beach, California, May 20, 1954.

expressed in stories to TAT cards in relation to socioeconomic status.

Hypotheses

It was the purpose of this study to explore specifically the correspondence between characteristics attributed to parents in TAT stories and in therapeutic interviews. The characteristics investigated were of two kinds: (*a*) attitudes toward the subject or patient which were ascribed to the parents; and (*b*) general descriptive images or pictures in which the subject has depicted the parental figures.

Using TAT stories and the content of therapeutic interviews of patients in a mental hygiene clinic, we attempted to test three hypotheses:

1. Patients who tell TAT stories involving attitudes ascribed to parents or images of parental figures will also in therapeutic interviews discuss attitudes which they ascribe to their own parents or images in which they are pictured. The converse also will be true. Hence, the TAT will furnish a basis for prediction.

2. Attitudes ascribed by patients to parental figures in TAT pictures will resemble those ascribed to their own parents (or parent substitutes) in therapeutic interviews conducted within a short period of time from the date of testing.

3. The kind of images in which parental figures in TAT pictures are described by patients will resemble images used in description of their own parents during these same therapeutic interviews.

It must be emphasized that in this study we were undertaking to discern only consciously expressed attitudes ascribed to the parents. No attempt was made to be subtle or interpretative in the notation of attitudes. Hence, if a description of a parental figure as "comforting and giving" indicated a concealed view of the parent as the exact opposite, this circumstance would not find inclusion in our data. Thus, possibly important aspects of TAT material such as reversals of conscious fantasy would not be revealed.

Method

The materials examined in this study came from the files of 50 patients in the Mental Hygiene Clinic of the Veterans Administration in Los Angeles. These materials were:

1. TAT stories in response to Cards 1, 2, 6 BM, 7 BM, 7 GF, and 18 GF—those cards in which figures appear which are the most easily identified or interpreted as parental ones; and

2. Records of ten therapeutic interviews which in most instances were the first ten interviews of therapy.

In order to avoid biased sampling as far as possible, the file of every patient tested in the year 1948 was examined consecutively and every file which met the criteria of containing the TAT and also the records of ten therapeutic interviews around the testing period was included. Since the year 1948 did not provide 50 cases, the files of the following year were examined in the same way until 50 such records were obtained. The 50 records used included 33 of males and 17 of females. It should also be noted that the specific cards used in the testing varied according to the judgment of the examiner; hence not all subjects were presented with each of the cards included in the study. Cards 1, 2, 6 BM, and 7 BM were used most frequently.

On the basis of a review of a sample of TAT stories and therapeutic interviews of 20 patients not used in the study, a check sheet was devised to include the 12 attitudes by which these patients characterized parental figures in the stories and in therapy sessions. The sheet provided a separate check list for each parent. Thus, it could be indicated that in the story in response to TAT Card 1, for example, a dominating attitude was attributed to the father, or that in response to 6 BM a solicitous attitude was attributed to the mother, etc. An identical check sheet was used to note the attitude mentioned during therapy sessions, and on this sheet the interviews (from 1 to 10) were indicated.

The attitudes included on the check sheets were as follows:

1. Comforting, giving, affectionate
2. Solicitous, concerned
3. Advice giving
4. Indifferent
5. Excessively attached
6. Over-giving
7. Dependent
8. Self-centered

9. Demanding (includes nagging; to get things for the parent)
10. Dominating, authoritative, wants to control, direct
11. Scolding, punishing
12. Rejecting, critical, obstructing, disapproving.

In the same manner a check list of images in which parental figures were depicted was also designed. These were:

1. Attractive, good, appetizing
2. Competent, effective
3. Incompetent, ineffective
4. Bad, evil, unattractive
5. Friction between parents.

Thus the check list included a total of 17 items.

The TAT stories and the ten therapeutic interviews were then analyzed by the two investigators together.[2] Every time that the presence of an attitude or an image in stories for each TAT card or in each therapy interview was noted, the item was marked on the pertinent check list in relation to the parent to whom it applied. If no attitude or image was discernible in a story on a TAT card or in an interview, this fact was indicated. The TAT stories and therapeutic interviews were analyzed independently, without knowledge of which stories were associated with which interviews.

Results

The data were analyzed to answer the three questions raised.

1. Do TAT stories serve as a basis for prediction as to whether or not patients who discuss parental attitudes and images in the test will discuss them also in therapy sessions? A chi-square analysis yielded no significant results in regard to:

Fathers' attitudes
Mothers' attitudes
Fathers' images
Mothers' images.

The results indicated clearly that the TAT furnished no basis for prediction in this situation.

[2] A sample of ten cases involving 330 judgments was analyzed by the investigators separately in order to obtain a measure of reliability. Phi coefficients of reliability were .54 for TAT protocols and .84 for the therapy sessions. Both of these values are significant considerably beyond the .001 level.

2. Do the attitudes ascribed to each parent in the TAT resemble those expressed in the therapy sessions? Analysis[3] of the data gave the following results:

1. In relation to fathers' attitudes, only six records out of the 50 had any item in common in the TAT and therapy; none had more than one item. Of the six records, only three could be interpreted as challenging the null hypothesis at a level of confidence between 5 and 10 per cent.

2. In regard to mothers' attitudes, 15 of the 50 records had one item in common and one had two in common. Only the latter could be interpreted as challenging the null hypothesis at a level of confidence between 5 and 10 per cent.

Hence, the results indicated that the kind of attitudes as related to parents in TAT stories furnished no basis for prediction of those that were ascribed to them in therapy sessions.

3. Do the images of parents as noted in TAT stories correspond to those expressed in therapy? Analysis of the data in answer to this question gave the following results:

1. In relation to fathers' images, only five records out of the 50 had any item in common, only one of which was significant at the 10 per cent level.

2. Of mothers' images, only five had any item in common, none more than one, and none reached even a 10 per cent level of significance.

Thus, analysis of the data bearing on all three of our inquiries corroborated a null hypothesis of little or no association between material in the TAT and material in the therapeutic interviews.

It seemed possible that one explanation of this low degree of association might result from the fact that the attitudes and images itemized on the check list were not sufficiently exclusive to permit precise differentiation by the investigators. To explore this possibility, the items on the check lists were divided arbitrarily into two groups, one group containing "positive" or favorable attitudes and images, the other "negative" or unfavorable ones.

[3] A combinatorial analysis employing the following formula, suggested by Dr. Joseph Gengerelli, was applied to the data:

$$\left[\frac{C^{M-(N_c+N_b)}_{N_a}}{C^{M}_{N_c+N_a}}\right]\cdot\left[\frac{C^{M-(N_c+N_a)}_{N_b}}{C^{M}_{N_c+N_b}}\right]\cdot\left[C^{M}_{N_c}\right]$$

where M = total number of possible choices, N_a = unique choices in TAT, N_b = unique choices in therapy, and N_c = choices in common.

This procedure, however, resulted in only a negligible increase in predictive value.

In the course of this dichotomized type of analysis, an interesting trend was noted. It was observed that some patients ascribed both positive and negative attitudes to the same parent. For example, a father would be described as "comforting and giving" at one point, and as "rejecting" or "dominating" at another. This "ambivalence," as we called it, occurred both in TAT stories and in therapeutic interviews in relation to both parents. However, in both the TAT and in therapy sessions the ambivalence expressed in regard to mother figures was significantly[4] greater than that in relation to fathers. This finding is consistent with many writings which have suggested that in our culture the relationship with the mother is the more complex and the more important in emotional illness.

Discussion

In interpreting the negative findings reported in this study, two considerations should be taken into account.

In the first place, it has already been mentioned that not all of the TAT cards selected for study were administered to each patient. The incompleteness of TAT data reduced the opportunity for producing material. Had additional material been available, it might have increased the value of the TAT in the predictions discussed.

In the second place, there is a possibility that TAT pictures evoke descriptions of parental attitudes which are not conscious to the patient at the beginning of therapy, but which become conscious to him in the course of therapy. If this possibility is a valid consideration, then little or no relationship between attitudes expressed in TAT stories and in the early stages of therapy should be expected. However, then we would predict that the attitudes expressed at some later point in successful insight therapy would correspond more closely to those in the TAT stories. A follow-up study investigating this possibility would be desirable.

[4] Critical ratio of difference between proportions in TAT stories, 2.4; in therapy sessions, 3.2.

Summary

In this study relationships between parental attitudes and images expressed by patients to TAT cards and in psychotherapy were investigated. Findings suggest no predictability from TAT to therapy as to whether or not parents were discussed. When they were discussed, no similarity was found between those attitudes and images given in TAT stories and in psychotherapy. Ambivalent attitudes toward both parents were present, but significantly more often toward mothers.

Received June 11, 1954.

References

1. Bell, J. E. *Projective techniques.* New York: Longmans, Green, 1948.
2. Bellak, L. The Thematic Apperception Test in clinical use. In L. E. Abt & L. Bellak (Eds.), *Projective psychology.* New York: Knopf, 1950. Pp. 185–229.
3. Butler, Pearl. Parent figures in TAT records of children in disparate family situations. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1948.
4. Combs, A. W. The use of personal experience in the Thematic Apperception Test story plots. *J. clin. Psychol.,* 1946, **4**, 358–363.
5. Harrison, R. Studies in the use and validity of the Thematic Apperception Test with mentally disordered patients. II. A quantitative validity study. III. Validation by the method of "blind analysis." *Charact. & Pers.,* 1940, **9**, 122–138.
6. Markmann, R. Predictions of manifest personality trends by a thematic analysis of three pictures of the Thematic Apperception Test. Unpublished thesis, Radcliffe College, 1943.
7. Mitchell, H. E. Social class and race as factors affecting the role of the family in Thematic Apperception Test stories of males. Unpublished doctor's dissertation, Univer. of Pennsylvania, 1950.
8. Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.
9. Murray, H. A. *Thematic Apperception Test. Manual.* Cambridge, Mass.: Harvard Univer. Press, 1943.
10. Rotter, J. B. Thematic Apperception Tests: suggestions for administration and interpretation. *J. Pers.,* 1946, **15**, 70–92.
11. Singer, J. L. Projected familial attitudes as a function of socioeconomic status and psychopathology. *J. consult. Psychol.,* 1954, **8**, 99–104.
12. Tomkins, S. S. *The Thematic Apperception Test.* New York: Grune & Stratton, 1947.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Application of a Psychometric Scoring Procedure to a Group Modification of the Thematic Apperception Test (N-TAT)[1]

Leonard D. Eron, Florence Sultan, and Frank Auld, Jr.
Yale University

This paper reports an extension of a psychometric scoring procedure (3), which was worked out with the standard TAT method, to a group TAT used by the Navy. The Navy TAT (N-TAT) has been described by Briggs (1) in a previous report.

Subjects and Procedure

The subjects for this study were 100 sailors chosen systematically from two classes at the U. S. Navy Submarine School at New London, Connecticut. The N-TAT protocol of every fifth man in these classes was drawn from the files until a sample of 100 had been obtained. Because the method of selection appears to be unbiased, we can assume that the group of 100 is representative of all submarine school entrants.

The N-TAT protocol consists of 10 stories written in a booklet on each page of which is a different stimulus picture. Seven of the pictures are slight modifications of the standard third edition series (8) and three are original.[2] For the purposes of this study, after each page of each booklet had been assigned an identifying number, the booklets were taken apart so that all stories for any given picture could be analyzed at the same time.

Each story was analyzed for the following variables: emotional tone, outcome, themes, perceptual distortions, unusual formal characteristics, and level of interpretation. All the stories were analyzed for a single variable at a time and scoring on one variable was completed for 100 stories before proceeding to the scoring of the next variable. Once such an analysis had been made, it was possible to compile both apperceptive and thematic norms (cf. Rosenzweig, 10), since the same scheme can be used for each picture and the scores on one picture are comparable to the scores on each of the other pictures. The reliability of the scoring scheme was tested by having two judges (Eron and Sultan) independently score 20 stories for each picture (a total of 200 stories for all 10 pictures). Results of the reliability study are reported below in connection with the discussion of each variable.

Results and Discussion

Emotional tone. In the analysis of the emotional tone of the stories, a set of five-point scales empirically derived for the standard set of pictures was utilized (6). Since three of the pictures are original with the N-TAT, the specific scales could be used for only seven and the general rating scale for emotional tone was used for the others. A Pearson r of .89 between two raters on 200 stories

[1] This study was supported by the Office of Naval Research, through the American Council on Education, and by the Foundations Fund for Research in Psychiatry. The conclusions reported here do not necessarily reflect the official opinion of the U. S. Navy. Thanks are extended to Commander G. J. Duffner of the U. S. Naval Submarine Base for his cooperation and support.

[2] The actual plates were reproduced in the earlier article (1). They are identified for our purposes as follows: 1. "The Operation," 8 BM in the 3rd edition. 2. "10 Minute Break," 9 BM in the 3rd edition. 3. "Father-Son," 7 BM in the 3rd edition. 4. "Captain's Mast," original. 5. "Submarine in Storm," adapted from 19 in the 3rd edition. 6. "Morning After," 13 MF in the 3rd edition. 7. "Contemplation," adapted from 14 in the 3rd edition. 8. "Pickup," original. 9. "Persuasion," 4 in the 3rd edition. 10. "Seabag," original.

Table 1

Percentage Frequency of Stories of Various Emotional-Tone Categories for Each of the TAT Pictures in the Navy Series

Picture no.	Emotional-tone rating					
	−2	−1	0	+1	+2	Reject
1	10(13.3)*	79(40.0)	5(9.3)	6(37.3)		
2	2(0.0)	24(26.7)	41(48.0)	33(24.0)	0.0(1.3)	
3	2(5.3)	49(58.7)	33(32.0)	14(4.0)		2(0.0)
4	1	92	2	5		
5	3(9.3)	61(33.3)	23(14.7)	3(33.3)	0.0(6.7)	10(2.7)
6	12(45.3)	70(49.3)	11(5.3)	6(0.0)		1(0.0)
7	8(9.3)	70(18.7)	10(45.3)	12(22.7)	0.0(4.0)	
8		61	32	6		1
9	4(4.0)	81(94.7)	7(1.3)	6(0.0)		2(0.0)
10		6	17	73	3	1

* Figures in parentheses refer to percentage frequency in original standardization group (3).

compares favorably with the reliability estimates obtained in other investigations using these scales with male and female college students ($r = .86$) (6) and adolescent girls ($r = .93$) (7). Table 1 presents the emotional-tone ratings for stories written by 100 sailors in response to each of the 10 N-TAT pictures.

For purposes of comparison, the percentage of ratings in each emotional-tone category in stories contributed in response to the corresponding pictures by 75 college student subjects of similar age and sex to whom the Murray cards were individually administered in the standard manner is included in parentheses. There are marked and statistically significant differences (beyond the .001 level of confidence) in the distribution of emotional-tone ratings between the two groups of subjects on three of the pictures. On picture 1 (The Operation) the college students contribute more +1 and fewer −1 stories. This no doubt occurs because the college group has more stories of aspiration toward a medical career and the sailor group has more stories about casualties at sea. For picture 5 (Submarine in Storm) there is the same kind of difference. Since this picture is a definite alteration of 19 in the standard series, designed to make the stimulus look like a submarine in distress, the sailors have more of this kind of story than of the usual story about the warmth and comfort of home in the midst of storms, which appears among the college students for that picture. The story of the comforts of home is, of course, a more positively toned interpretation than the one about a submarine in distress. On picture 6 (Morning After) the sailors contribute more −1, and the college students more −2 stories. That the college group gives sadder stories for this picture perhaps reflects the fact that in this group the theme of death of the wife is equally as prominent as the theme of remorse after intercourse, whereas among the sailors the latter is the modal theme. This may reflect a personality difference in the two groups. For picture 7 (Contemplation) the sailors have more −1 stories and the college group more 0 stories. This may be a function of the alteration in the picture—from a figure looking out of a window to a sailor leaning over a rail.

Table 2 presents thematic norms for emotional tone—the percentage of subjects who have a given number of stories of each emotional-tone rating in a series of 10 stories. There is no real basis for comparison with thematic norms previously compiled in other groups, since they have been based on all 20 pictures in the standard series. However, it is obvious that, just as in the individually ad-

Table 2

Percentage Frequency of Individuals Who Have a Given Number of Stories of Each Emotional-Tone Category in a Total of Ten Stories Given in Response to the Navy TAT

No. of stories	Emotional-tone rating							
	−2	−1	0	+1	+2	Reject	*X*	*?*
0	67		20	9	97	93	99	94
1	26	2	27	49	3	6	1	4
2	5		27	22		1		2
3	2	2	11	12				
4		12	11	5				
5		18	2	2				
6		31		1				
7		26	1					
8		6	1					
9		3						

ministered TAT, these pictures too elicit predominantly sad stories.

Outcome. A set of five-point scales empirically derived on the standard series of pictures in another investigation (3) was used in the analysis of outcomes. The inter-interpreter reliability derived from the independent ratings of 200 randomly selected stories by two judges was .81 (Pearson r), which compares favorably with reliabilities obtained in three other studies (3, 4, 7). It was necessary to add an additional category, *X*, to the scale to cover those stories in which there was no mention of any outcome. In the individual administration it is always possible to ask for an outcome when the subject fails to give one. This is, of course, impossible in a group administration. Thus only when the subject gives alternative outcomes or specifically writes that he cannot make one up is he scored with a *?* outcome in this procedure. Otherwise, if there is no definite outcome, he is scored *X*. Consequently, in this group of subjects the largest category for any story was *X*, and a comparison with the standardization group is not valid. However, the outcome ratings of all stories contributed by the Navy group for each of the pictures are provided in Table 3 as normative data to be consulted when the N-TAT or a similar kind of adaptation is used. Similarly, thematic norms (for a series of 10 pictures) are provided in Table 4.

Themes. In analyzing the themes, the second edition of the check list, the derivation of which has been described elsewhere (2), was used. Since there are three new pictures (Captain's Mast, Streetcorner Pickup, and Seabag) in this series, it was necessary to revise the check list further and add some new themes.[3] However, the number of changes necessary was not significant. Reliability of the "Navy" check list was good. Percentage agreement between two independent raters was calculated according to the following formula:

$$\frac{(\text{No. of agreements between Judge A and Judge B}) \times 2}{\text{Total judgments A} + \text{Total judgments B}}$$

It was 89 per cent for a total of 200 stories (20 stories from each of 10 pictures). This compares well with what has been obtained in another study by different investigators (7).

Table 3

Percentage Frequency of Stories of Various Outcome Categories for Each of the TAT Pictures in the Navy Series

Picture no.	Outcome rating							
	−2	−1	0	+1	+2	Reject	*X*	*?*
1	10	6	1	21	2		49	11
2	1	18	26	13	1		37	4
3	2	10	8	19	1	1	54	6
4	1	42		19	2		26	10
5	3	5	7	15	2	8	57	3
6	9	23	9	7	2	1	35	14
7	5	8	8	21			53	5
8		21	11	13		1	41	13
9	3	12	5	19		2	54	6
10		4	6	44	4	1	33	8

Table 4

Percentage Frequency of Individuals Who Have a Given Number of Stories of Each Outcome Category in a Total of Ten Stories Given in Response to the Navy TAT

No. of stories	Outcome rating							
	−2	−1	0	+1	+2	Reject	*X*	*?*
0	74	26	49	21	92	91	15	47
1	19	30	31	33	5	5	9	32
2	6	24	14	17		3	10	15
3	1	12	5	12	3	1	6	6
4		6		7			11	
5		1		5			8	
6			1	5			13	
7		1					12	
8							8	
9							5	
10							3	

The most frequent themes for each picture contributed by 10 per cent or more of the subjects in this group of 100 sailors or in the standardization group (for the comparable pictures only) are listed

[3] The Navy check list of themes, defining each theme and noting the changes, deletions, and additions to the second edition, has been deposited with the American Documentation Institute. Order Document No. 4421 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

in Table 5. The chief differences between the two groups were alluded to in the discussion of emotional tone. They appear to be largely a function of the physical differences between the original Murray pictures and the adaptations used in this study, as well as of the different life situations and predominant needs of the subjects. Because of the two variables—the stimuli and the subjects—which are confounded, it is impossible to partial out the exact cause of the differences in results. However, those who use this or similar adaptations of the TAT with similar subjects should be aware of these differences when making interpretations. Thematic norms appear in Table 6.

Levels of interpretation.[4] It was necessary to develop a new scale for levels of interpretation in this study because of the difference in administration from the standard method. Stories were designated on an eight-point scale going from complete rejection of the picture to complete narrative (including a

[4] This category does not refer to a variable which carries a similar name in previous publications (2, 3, 4). The latter variable is now called "unusual formal characteristics."

Table 5

Most Frequent Themes in Stories Given in Response to Each of the TAT Pictures in the Navy Series*

Picture		Theme	Percentage frequency	
			Navy	College
1				
(The Operation)	IA3i	Death or illness of peer	31	2.7†
	IB12	Death or illness of central character	25	14.7
	IC6	War	21	29.3
	IB29	Death, illness—unspecified character	18	
	IB14	Reminiscence, sad	17	6.7
	IB7	Guilt	15	
	IC5	Aggression from—impersonal	12	8.0
	IA3g	Concern—peer	11	
	IB1	Aspiration	6	53.3†
	IA1a	Death, illness of parent	7	18.7†
2				
(10 Minute Break)	IB4	Retirement	47	68.0†
	IB21	Exhaustion	45	14.7†
	IIB3	Reminiscence, happy	31	4.0†
	IC7	Escape from perilous environment	13	
	IC6	War	10	9.3
	IB19	Vacillation		13.3
	IIB2	Tranquillity		16.0
3				
(Father-Son)	IAib	Succorance from parent	62	74.7†
	IA11	Disappointment in parent	14	
	IA1a	Pressure from parents	6	28.0†
	IB11	Occupational concern	4	25.3†
4				
(Captain's Mast)	IA1a	Pressure from parent	92	
	IC7	Escape from perilous environment	35	
	IB13	Retribution	31	
	IC4	Aggression toward environment	25	
	IB7	Guilt	17	
5				
(Submarine in Storm)	IC5	Aggression from—impersonal	42	61.3†
	IC6	War	26	5.3†
	IIA1f	Contentment at home		26.7
	IB9	Fear		10.7
6				
(Morning After)	IA2h	Illicit sex	64	42.7†
	IB7	Guilt	35	54.7†
	IB8	Drunkenness	17	13.3
	IA2j	Death, illness of partner, female	15	42.7†
	IA2v	Venereal disease, partner	14	
	IA2e	Aggression toward partner		17.3
	IB13	Retribution		16.0

Table 5—*Continued*

Picture		Theme	Percentage frequency	
			Navy	College
7				
(Contemplation)	IIB3	Reminiscence, happy	27	12.0†
	IB28	Homesickness	22	
	IB17	Loneliness	11	5.3
	IA2s	Unrequited male partner	10	
	IB5	Suicide	10	2.7
	IB3	Curiosity		26.7
	IB1	Aspiration		22.7
	IB4	Behavior disorder		14.7
	IB10	Rumination		12.0
	IIC1	Favorable environment	5	10.7
	IIB2	Tranquillity		10.7
8				
(Pickup)	IIB7	Ordinary activity	34	
	IA2pp	Decision, female partner	23	
	IA2n	Competition, partner	21	
	IA2q	Pursuit, partner	17	
	IA1a	Pressure, parent	15	
	IA11	Disappointment in, parent	15	
	IC2	Legal restriction	11	
	IA1b	Succorance, parent	10	
9				
(Persuasion)	IA2r	Seduction	24	10.7†
	IA2c	Succorance, partner	22	22.7
	IA2a	Pressure, partner	19	50.7†
	IA2xx	Ulterior motive	14	
	IA2f	Departure from partner, female	13	16.8
	IB12	Illness or death central character	13	
	IA2h	Illicit sex	9	12.0
	IA2m	Competition		12.0
	IB6	Moral struggle		12.0
10				
(Seabag)	IIA1d	Reunion, parent	51	
	IIB3	Reminiscence, happy	27	
	IB1	Aspiration	13	
	IIB7	Ordinary activity	13	
	IIB5	Occupational satisfaction	12	

* Only those themes contributed by 10 per cent or more of the subjects in either group are included in this table.
† Themes for which the difference between groups is significant beyond the .05 level of confidence.

conclusion and interpretation of feelings and thoughts of the characters). The scale appears in Table 7. The percentage agreement between two raters for 200 stories (20 stories for each of 10 pictures) was 85. The incidence of the various levels for each story appears in Table 8. From this table and Table 9, which presents thematic norms, it can be seen that, although certain of the pictures tend to elicit less well-developed stories, the general level of most of the responses indicates a good deal of involvement on the part of the majority of subjects, even in this group method of administration.

Perceptual distortions. From the number of perceptual distortions listed in Table 10 it can be seen that perceptual distortions are not nearly so frequent among the group of sailors as they were among the original standardization group (3). The lack of distortions is probably due to group administration, because the subjects can study the pictures more carefully and need not write down their first impressions. Also, since a new set of pictures was drawn up for this series, it may be that these pictures differ in ambiguity from the original set. One outstanding difference in results between this group and previous groups is in the number of rejections. In the previous studies such behavior was rare even among subjects who were psychiatrically ill (2, 3); in the Navy group, 10 per cent of the

Table 6

Percentage Frequency of Individuals Who Have a Given Number of Specific Themes in a Total of Ten Stories*

No. of theme	Theme																	
	IA1a†	IA1b	IA1g	IA1i	IA1d	IA2g	IA2h	IA2k	IA2q	IA2r	IA2u	IA2v‡	IA2xx‡	IA3b	IA3c	IA3i	IA3j	IA3l
0	9	29	96	91	74	89	32	92	82	68	92	81	83	93	89	67	93	94
1	62	54	3	8	24	9	55	7	13	30	7	17	15	6	10	31	6	5
2	25	15	1	1	2	2	10	1	4	2	1	1	2	1	1	2	1	1
3	3	2					2		1									
4	1						1					1						

No. of theme	Theme																
	IB1	IB2	IB3	IB4	IB5	IB6	IB7	IB8	IB9	IB10	IB11	IB12	IB13	IB14	IB16	IB17	IB21‡	IB34‡
0	72	89	91	85	89	95	47	61	72	95	85	59	65	74	89	88	53	85
1	22	9	5	14	10	4	33	32	25	4	14	36	33	18	10	11	43	14
2	5	2	4	1	1	1	18	4	3	1	1	5	2	6		1	4	1
3	1						2	3						2	1			
4																		

No. of theme	Theme													
	IC2‡	IC3	IC4‡	IC5	IC6	IC7‡	IC8	IIA1d	IIA3d	IIA2e	IIB3	IIB5	IIB7	IIC1
0	81	94	70	45	57	52	93	51	91	92	40	86	50	89
1	17	5	28	50	27	39	6	44	8	7	40	13	39	9
2	2	1	2	5	7	9	1	5	1	1	15	1	10	2
3					5						4		1	
4					4						1			

* Only those themes contributed at least twice by any subject are included in this table.
† Themes given here by code number have been defined in Table 1 in a previous article published in this journal (1953, 17 [4], 269–282). The themes indicated by ‡ are original with the N–TAT and thus do not appear in that list. They are: IA2v, venereal disease; IA2xx, ulterior motive; IB21, exhaustion; IB34, grief; IC2, legal restriction; IC4, aggression toward environment; IC7, escape from perilous environment. Definitions of these themes appear in the Navy check list deposited with the American Documentation Institute (see footnote 3).

subjects rejected at least one picture. The increased rejection rate is probably a function of the group administration method. It is easier to insure maximum cooperation throughout an entire test series when the experimenter is present to prompt and encourage the subject throughout as in the individual administration. With the present method it is difficult to assess the exact cause of this large number of rejections.

Application of the Method

The practical utility and validity of this type of psychometric analysis of the N-TAT

Table 7

Level of Interpretation

A	Rejection
B	Description of physical appearance of objects and/or persons
C	Description of action
D	Description of feelings, attitudes, and/or thoughts of characters
E	Incomplete narrative—omits either antecedents, present action, or conclusion
F	Complete narrative—includes antecedents, present action, and conclusion
G	Incomplete narrative with interpretation of feelings, attitudes, and/or thoughts
H	Complete narrative with interpretation of feelings, attitudes, and/or thoughts
W	Not really a story, reflection on life, words, advice, remonstrance

Table 8

Percentage Frequency of Stories of Various Level-of-Interpretation Categories for Each of the TAT Pictures in the Navy Series

Picture no.	Level rating								
	A	B	C	D	E	F	G	H	W
1			6	8	2	11	33	35	5
2			7	5	12	27	12	32	5
3	1		10	2	27	16	12	29	3
4			4	3	12	30	7	40	4
5	8	15	24		10	27	5	6	5
6	1		7	6	7	6	16	50	7
7			5	18	3	5	24	39	6
8	1		8		18	26	12	30	5
9	2		13	2	9	11	28	31	4
10	1		13	8	2	18	15	35	8

Table 9

Percentage Frequency of Individuals Who Have a Given Number of Stories of Each Level-of-Interpretation Category in a Total of Ten Stories Given in Response to the Navy TAT

No. of stories	Level rating								
	A	B	C	D	E	F	G	H	W
0	91	85	60	61	41	28	33	12	82
1	5	15	16	28	30	26	23	21	9
2	3		12	9	16	17	20	14	3
3	1		4	2	11	15	7	14	2
4			2		1	7	8	5	1
5			1		1	3	7	11	
6			4			2	2	8	
7						2		9	
8			1					3	1
9								2	1
10								1	1

were demonstrated in a study the results of which are fully reported elsewhere (5). A comparison was made of TAT stories told by 22 sailors before they were sealed up in a submarine and the stories told by them later, while they were still enclosed. It was possible to demonstrate, on the basis of the protocols, in what ways the 22 men who volunteered for the experiment ("Operation Hideout") differed from the usual run of submariners, and also to enumerate the differences which appeared in their fantasy productions after the experimental treatment. These changes in fantasy coincided with differences observed in physical, physiological, and psychomotor functions.

Table 10

Percentage Frequency of Perceptual Distortions for Each of the TAT Pictures in the Navy Series

Picture no.	Distortion	Frequency
1	Patient as female	13
	Picture within picture	1
2	Responds to only one character	2
3	Self-image	1
4		
5	Stove or gun cover	1
	Serpent	1
	Fish	3
	Rabbit	1
	Locomotive—cat face	1
	Dragon	1
6		
7		
8		
9	Female seen as mother	1
10	Gun with arms and legs	1

Summary and Conclusions

A psychometric method of analysis of TAT stories derived from protocols obtained in the standard manner was extended to a group adaptation of the TAT which is used by the U. S. Navy. The test records of 100 systematically selected submariners were analyzed for emotional tone, outcome, themes, level of interpretation, and perceptual distortions. The reliability of the scales and check lists used was demonstrated to be of sufficiently high order to warrant their use with this type of administration and this type of subject. Both apperceptive and thematic norms are provided for the convenience of those using this adaptation. This psychometric procedure was tried and found to be useful and valid in a research which is reported in a separate monograph.

Received June 16, 1954.

References

1. Briggs, D. L. A modification of the Thematic Apperception Test for naval enlisted personnel (N-TAT). *J. Psychol.*, 1954, **37**, 233–241.
2. Eron, L. D. Frequencies of themes and identifications in the stories of schizophrenic patients and non-hospitalized college students. *J. consult. Psychol.*, 1948, **12**, 387–395.
3. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
4. Eron, L. D. Responses of women to the Thematic Apperception Test. *J. consult. Psychol.*, 1953, **17**, 269–282.
5. Eron, L. D., Auld, F., Jr., Lesser, G., Laffal, J., & Sultan, Florence. A study of the TAT stories and sentence completions of subjects in "Operation Hideout." *Bur. Med. Surg., Navy Dept., M.R.L. Rep.*, 1954, No. 243.
6. Eron, L. D., Terry, Dorothy, & Callahan, R. The use of rating scales for emotional tone of TAT stories. *J. consult. Psychol.*, 1950, **14**, 473–478.
7. Liccione, J. V. The changing family relationships of adolescent girls. Unpublished doctor's dissertation, Syracuse Univer., 1954.
8. Murray, H. A. *Manual for the Thematic Apperception Test.* Cambridge, Mass.: Harvard Univer. Press, 1943.
9. Ritter, Anne M., & Eron, L. D. The use of the Thematic Apperception Test to differentiate normal from abnormal groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 147–158.
10. Rosenzweig, S. Apperceptive norms for the Thematic Apperception Test. I. The problem of norms in projective methods. *J. Pers.*, 1949, **17**, 475–482.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Effects of Sexual Stimulation Under Varying Conditions on TAT Sexual Responsiveness[1]

Paul H. Mussen and Alvin Scodel
The Ohio State University

The present study is concerned with the effects of sexual stimulation in different situational contexts on subsequent TAT sexual responsiveness. Our hypothesis states that these responses will be more inhibited following sex arousal in the presence of a stern father figure than following such arousal in the presence of a young, permissive individual.

A procedure almost identical with the one described by Clark (1) was used. The subjects were two groups of male students in an introductory psychology course. Both groups took the TAT after having been exposed to sex-arousing stimulation. In the Formal group (33 subjects), a series of eight photographic slides of attractive nude females was presented by a formal, professorial, and somewhat stern man in his sixties. Subjects were told that the research was related to Sheldon's body-type theories. They were instructed to rate the attractiveness of each of the eight nudes and to indicate the criteria for their judgments.

Using the same instructions, a young-looking, informal, permissive graduate student presented the slides to the Informal group (38 subjects).

[1] An extended report of this study may be obtained without charge from Paul H. Mussen or Alvin Scodel, Ohio State University, Columbus 10, Ohio, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4439 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

In both groups, the first examiner left the room immediately after the ratings were completed. At this point, another experimenter stepped forward, explained that he was attempting to standardize a test of creative imagination, and then administered eight TAT cards from the Murray series (7BM, 12BG, 3BM, 14, 20, 6GF, 6BM, 13MF) to the group.

The manifest and implied sexual content of each TAT story was scored according to a seven-point scale, ranging from 0 to 6, depending on the intensity of sexual activity. The subject's total sexual response score was the sum of the scores on the eight stories.

The mean sexual response score of the Informal group was 9.55, while the mean for the Formal group was 7.12 (*SD*'s, 5.18 and 4.26, respectively). The difference between means yielded a t value of 2.13, significant between the 2 and 5 per cent levels of confidence.

Thus, it would seem that arousal of socially disapproved needs in the presence of an authority figure results in greater inhibition of subsequent expression of that need than arousal of the same need in a permissive situation.

Brief Report
Received November 11, 1954.

Reference

1. Clark, R. A. Projective measurement of sexual motivation. *J. exp. Psychol.*, 1952, **44**, 391–399.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Diagnostic Significance of Usual and Unusual TAT Stories

Jerome S. Miller and Alvin Scodel

The Ohio State University

In spite of numerous efforts, direct and frontal attempts to validate the TAT have not been convincingly successful. Contributing to this difficulty are such factors as variations in administration and scoring procedures employed by users of the TAT and differences in interpretation as a function of the clinician's orientation and purpose. But, perhaps most important, the major problem in validation of the TAT lies in the elusiveness of adequate criterion data, combined with a lack of clinical or theoretical agreement as to just what it is that is to be validated.

Despite these obstacles to its demonstrable acceptability, other approaches are available for determining the TAT's efficiency for eliciting meaningful interpretive data. For example, construct validity, or the principles upon which such an instrument is based, often can be put to test. Demonstrating the plausibility of underlying assumptions is another way which also can lend to the credibility of inferences from such an instrument.

Lindzey has recently published an article exploring the latter of these two alternatives (4). He examined a number of pertinent assumptions underlying the TAT and evaluated them in the light of empirical evidence available in the literature. Our present study was designed to extend Lindzey's alternative method by isolating one underlying assumption and testing it within a TAT situation.

The specific assumption to be tested holds that stimulus-bound responses are not as diagnostically significant as responses which are less constrained by the stimulus material. Lindzey stated that this assumption—" . . . Themes or Story Elements That Appear to Arise Directly out of the Stimulus Material [Are] Less Significant than Those That Are Not Directly Determined by the Stimulus Material . . . "—is consistent with data from the investigations of "new look" perceptionists (4, pp. 13 f.).

Stated in another way, the assumption we studied refers to the significance of the unusualness of a response, i.e., the degree to which an idiosyncratic or atypical fantasy production is diagnostically useful. One recent study states in the introduction that, "Despite the scarcity of normative data, various clinical investigators have stated that the presence of certain characteristics or 'deviant' content in a TAT record provides a basis for differential diagnosis" (8, p. 147). The same article concludes " . . . in the interpretation of a given protocol it is felt that this study offers further evidence that fantasy material is important chiefly when it deviates from the expected response" (p. 157).

Rotter states that "Occasionally the subject tells a story which has little or no significance in terms of his basic problems. Faced with a stimulus which he does not react to emotionally, he gives some common interpretation based on relatively superficial experiences with such situations" (9, pp. 83–84). Murray makes a somewhat similar point when he emphasizes the interpreter's task of " . . . noting everything that is unusual, uncommon or unique; or common but unusually high or low in intensity or frequency (5, pp. 7–8). Or as put by Rapaport *et al.*, "The standard cliché is always approximated but never reached; it is an abstraction from which deviations may be measured. The greater the deviation, the more amenable is the story to diagnostic interpretation" (7, p.

421). In general then, many psychologists interpret TAT themes on the basis of variation from implicit or explicit norms of typical themes.

Method

Normative data for the TAT have been notably lacking. Therefore, there are no available standards by which to measure the degree to which a given response is determined by the stimulus material. Moreover, there are no empirical norms representing the common or usual response elements for each stimulus. An exception to this deficiency exists in the introductory set of card-by-card norms recently published by Eron (1). These norms encompass a rating scale for emotional tone of TAT stories; a frequency count of descriptive theme categories and interpretation levels; outcome ratings of stories; frequency of direction of shift in emotional tone from the body of the story to the outcome; identification of characters and objects; and the extent to which subjects single out some unusual or minute aspect of the picture for elaboration in their stories. The data were collected from a total sample of 150 male veterans of World War II.

An earlier attempt by the present writers to test the hypothesis of the diagnostic significance of thematic unusualness, using Eron's normative data, was unsuccessful. The plan, in brief, was to obtain instances where two individuals had each given a story subsumed under a single category except that each instance was in response to different stimulus cards. One production was considered atypical in terms of the normative frequency data and the other story was considered as typical. However, several attempts to obtain reliability between judges in labeling the thematic categories of stories resulted in only 50 per cent agreement and the plan was therefore rejected.

Selection of the story population. For the procedure used in the present study 35 sets of TAT stories were obtained. Each set contained three stories given to each of five TAT stimulus cards (1, 6BM, 7BM, 12M, and 16). Each subject's successive stories to any one card were given in response to instructions to "make up another story to this card, but make this story as different as possible from the story you made up before for the same picture." Subjects were male adults and the 35 represent all of such specially tested men who were in psychotherapy for two or more treatment sessions at a mental hygiene clinic. At the time of this study all 35 had discontinued treatment and were "closed" cases.

Table 1

Distribution of Sums of Squares of Two Judges Ranking Differences for a Five-Point Scale ($N = 35$)

Sums of differences squared	Frequencies		
	Theoretical	Observed	D^2/ Theoretical
0–10	7.87	29.0	56.73
12–16	5.84	2.0	2.53
18–22	7.59	4.0	1.70
24–28	5.84	0.0	5.84
30–40	7.87	0.0	7.87

Note.—$\chi^2 = 74.67$ ($df = 4$), $p < .0001$, $C = .72$.

These stories were presented to two experienced clinicians who independently ranked the stories on a five-point scale, the most unusual story in a protocol being ranked "1" and the least unusual (or most usual) story "5." Only the first stories given to each stimulus card were used. The two judges each employed their own private, implicit norms for this task.

Reliability of selection was measured by using Olds' form for the distribution of sums of squares of rank differences (6). A chi square was computed between this expected distribution and the distribution of D^2 between the two judges' rankings of the stories. This yielded a chi square of 74.67 with $df = 4$ and $p < .0001$ as shown in Table 1. Where the two selection judges were not in exact agreement at the extremes of the scale, a third experienced judge arbitrated the disagreements. In this way 35 usual and 35 unusual TAT stories were obtained.

Procedure. The 70 stories were typed on individual cards without further identifying data and presented to six new judges. The judges attempted "blind" matching to typescripts of the first two psychotherapy sessions of each of the 35 subjects. The cards and the psychotherapy records were presented to judges in groups. Each group consisted of either three or four sets of therapy records and the two stories—the usual and the un-

usual—corresponding to each of those records. The task was to match each of the stories with one of the records. The judges were unaware that there were two kinds of stories. They were aware, however, that the stories had been obtained originally so that it was possible for two different stories from the same stimulus card to have been given by one subject. The number of stories matched with any case record ranged from none to five.

Another precaution was to equate roughly for length all stories within any particular group. This was done to reduce the possibility of allowing productivity or verbality to serve as a basis for matching.

Results

The number of correct and incorrect matchings for the usual and unusual TAT stories are shown in Table 2.[1] Chi squares were computed for each judge separately. The over-all significance of combined results was also computed using the chi-square model of Jones and Fiske (3) on the assumption that each judge's results are unrelated to the probability associated with each of the other judges' results. This combined result provides a chi square of 12.46 which, with $df = 12$, gives a p value of .41 approximately (Table 2).

Table 2

Results of Judges in Matching Usual and Unusual TAT Stories with Psychotherapy Records

	Judge					
	A	B	C	D	E	F
Number of usual						
Matched	11	13	17	19	11	17
Unmatched	24	22	18	16	23	17
Number of unusual						
Matched	14	15	17	15	17	21
Unmatched	21	20	18	20	17	13
χ^2 ($df = 1$)	0.58	0.24	0.00	2.04	2.18	0.96
p	.46	.67	1.00	.15	.13	.33

Note.—Combined $\chi^2 = 12.46$ ($df = 12$), $p > .40$ (3).

Similarly, chi squares were computed after eliminating instances where a judge had correctly assigned both the usual and unusual stories (Table 3). This was done to determine whether or not some of the lack of difference might have been due to matching mediated by extraneous cues. A usual story, in some cases, might have been paired with its corresponding unusual story through similarity of language factors in the stories, administration techniques, the examiner's individual fashion of recording, etc., thus supervening direct matching to a therapy record. However, even with these cases deleted, the data fail to show a marked superiority for the matching of unusual stories.

Table 3

Results of Judges in Matching Usual and Unusual TAT Stories with Psychotherapy Records: Paired Matchings Eliminated

	Judge					
	A	B	C	D	E	F
Number of usual						
Matched	5	8	9	11	4	8
Unmatched	24	22	18	16	23	17
Number of unusual						
Matched	8	10	9	7	10	12
Unmatched	21	20	18	20	17	13
χ^2 ($df = 1$)	0.90	0.32	0.00	1.32	3.48	1.34
p	.35	.63	1.00	.26	.06	.26

Note.—Combined $\chi^2 = 14.04$ ($df = 12$), $p > .30$ (3).

To determine whether the frequency of correct matchings was high or low over all, a chi square was computed for usual and for unusual stories each versus chance expectation (Table 4). Chance frequency was 11 for each of the two kinds of stories for each judge. Three judges (C, E, F) matched unusual stories at better than the .05 level and three

Table 4

Judges' Matchings of Usual and Unusual TAT Stories: Each Versus Chance

	Judge					
Chi square	A	B	C	D	E	F
Usual vs. chance	0.00	.53	4.77*	8.49**	0.00	4.86*
Unusual vs. chance	1.20	2.12	4.77*	2.12	4.84*	13.44**

* Statistically significant at the 5% level.
** Statistically significant at the 1% level.

[1] Judges E and F matched against only 34 subjects since one individual resumed treatment and his data became unavailable. The chance probability of correct matching remained the same, however, for each judge.

(C, D, F) matched the usual stories at better than the .05 level. With the use of Wilkinson's table (10) the combined results show that both the usual and unusual stories were correctly matched more frequently than by chance at about the .002 level. Both usual and unusual stories then were apparently matched with better than chance accuracy, with neither type of story showing significant superiority.

Discussion

Essentially, our interest lay in a demonstration of a "back-door" approach to the problem of validating the TAT and similar fantasy-production techniques. Testing an underlying assumption is not, by any means, either a crucial test of the TAT's validity or a decisive illustration of such an alternative, back-door course. The surprising feature of the study is the lack of confirmation of an assumption generally accorded high face validity and wide subscription by clinicians.

Because of the questions provoked by the unexpected results, the data were further treated to discount the possibility of matching mediated by extraneous cues. To do this, those cases were excluded in which both the usual and the unusual stories were correctly identified with their common psychotherapy criterion. For the cases which remained, a statistically significant difference between the two kinds of stories still did not appear. In addition, as noted before, the number of successful matchings for each of the two kinds of stories exceeded chance expectations.

There are some speculative explanations of the accurate identification of usual stories with their corresponding psychotherapy records. One reason may be that stories labeled usual are unremarkable only in relation to other stories from the same protocol. Thus, our two populations of stories, differentiated by their relative ranks, actually may both have been loaded in the direction of atypicality.

However, the two selection judges showed high agreement in their independent rankings, and we must conclude that there were, in general, discernible differences between adjacent stories throughout the five ranks—not merely between the first and fifth ranks.

Secondly, no TAT card yields but one cliché response. For example, Eron found that stories to Card 1 fitted thematic categories of "aspiration," "parental pressure," and "belongingness" in 46.7, 41.3, and 25.3 per cent of his cases respectively (1). A judge or clinician might then derive a meaningful differentiating cue by noting which of several stereotyped responses is chosen by the subject. Also, even a hackneyed story can be suggestive by virtue of the intensity of some fairly common theme. Suppose, for example, that a subject has intense dependency needs. Then on Card 6BM, for example, we might expect that subject to depart only minimally, if at all, from a theme of dependency so commonly given in response to this card. Yet we might be able to differentiate this subject from another on the basis of how intensely his story reflects his need structure in this particular area. Thus, the cliché is also characteristic of a subject and may be revealing of him in a broad context. The presence of repeated stereotyped stories in a protocol may be indicative of a subject's method of handling his problem areas.[2]

In closing this explorative discussion, a final question is raised which concerns a difference between the task imposed upon the judges of this study and the procedure generally followed in a clinical interpretation. In the latter situation there is some pressure upon the worker to identify the cues and to integrate and verbalize hypotheses and diagnostic descriptions. None of these requirements existed for the blind matching procedure. It is conceivable that judges may have derived unverbalizable or subliminal cues from a story and matched them with similar unverbalizable or subliminal cues inferred from the psychotherapy typescripts. It is not known what effect this freedom from explicitness has upon the data. It is possible in a clinical setting, however, that the notion

[2] Holt remarks on this interpretative value of usual stories in his statement that "It may be said further that frequently a great deal more can be interpreted from a set of 'cliché stories,' than is frequently claimed. It is more difficult; it requires more experience and more sensitivity to subtle ways in which a stereotyped story differs from dead average; but . . . it is always possible to make some progress" (2, pp. 201–202).

of unusual stories having diagnostic advantage stems more from ease of interpretation than from demonstrably superior predictiveness.

Summary

In this study we attempted to test an assumption generally held to have significance in the interpretation of TAT stories and similar fantasy productions. The hypothesis tested was that judges using a blind matching technique will more often ascribe unusual TAT stories to corresponding psychotherapy records of subjects than they will correctly ascribe usual stories.

A usual story and an unusual story were reliably obtained from each of 35 psychotherapy clients. The results of the judges' matching attempts, however, did not confirm the hypothesis at a statistically significant level, although each type of story was matched significantly better than chance.

These results were briefly discussed and evaluated and it was tentatively concluded that the stereotyped, or cliché, TAT story is not without meaning characteristic of its producer and should not be overlooked as a resource for the clinician's understanding of a subject.

Received July 21, 1954.

References

1. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
2. Holt, R. R. The Thematic Apperception Test. In H. H. Anderson & G. L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 181–229.
3. Jones, L. V., & Fiske, D. W. Methods for testing the significance of combined results. *Psychol. Bull.*, 1953, **50**, 375–382.
4. Lindzey, G. Thematic Apperception Test: interpretive assumptions and related empirical evidence. *Psychol. Bull.*, 1952, **49**, 1–25.
5. Murray, H. A., *et al. Thematic Apperception Test. Manual*. Cambridge, Mass.: Harvard Univer. Press, 1943.
6. Olds, E. G. Distribution of sums of squares of rank differences for small numbers of individuals. *Ann. math. Statist.*, 1938, **9**, 133–148.
7. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing*. Vol. 2. Chicago: Year Book Publishers, 1946.
8. Ritter, A. M., & Eron, L. D. The use of the Thematic Apperception Test to differentiate normal from abnormal groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 147–158.
9. Rotter, J. B. Thematic Apperception Test: suggestions for administration and interpretation. *J. Pers.*, 1946, **15**, 70–92.
10. Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156–158.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Diagnostic Discrimination of the "Basic Rorschach Score"

Ben C. Finney
VA Hospital, Palo Alto [1]

The diagnostic adequacy of the Basic Rorschach Score (1) was tested by comparing the mean scores of groups with different diagnoses which were collected and scored by McReynolds (2, 3). The protocols were scored according to the "Diagnostic Sign List" (2, p. 80). In order to compare the results with those reported by Buhler *et al.* (1, Table III, p. 12), the 1945 and 1949 samples of comparable diagnostic groups of Buhler were combined on the basis of the statistics reported, and the neurotic divisions were combined into one single neurotic group. The statistics for Buhler's and McReynolds' samples are reported in Table 1.

In McReynolds' samples, there were no significant differences between the Neurotic vs. Normal or between the Paranoid vs. Mixed Schizophrenic groups. There were significant differences, at better than the .01 level, between the Normal vs. Paretic, Normal vs. Combined Schizophrenic, and Neurotic vs. Paretic groups. The difference between the Neurotic and the Combined Schizophrenic groups was significant at the .05 level.

When McReynolds' and Buhler's groups were compared, there were no significant differences between the Schizophrenic groups but Buhler's Normal group was significantly higher than McReynolds' Normals and Buhler's Neurotics were significantly higher than either McReynolds' Normals or Neurotics.

Table 1

Basic Rorschach Scores

Diagnostic group	*N*	Mean	*SD*
Normals (McR)	111	− 7.12	12.74
Neurotics, Mixed (McR)	22	−10.00	13.00
Schiz. Par. (McR)	31	−15.00	14.13
Schiz. Mixed (McR)	33	−18.76	10.14
Schiz. Combined (McR)	64	−16.94	12.38
Paretics (McR)	24	−26.42	9.13
Normal (Buh)	89	18.90	6.03
Neurotics, Comb. (Buh)	176	8.90	7.20
Schiz. Par. (Buh)	29	− 9.64	13.07
Schiz. Mixed (Buh)	63	−19.42	11.75
Schiz. Comb. (Buh)	92	−16.89	12.28

These results suggest that despite similar diagnoses, McReynolds' and Buhler's samples were not from a similar population and that within McReynolds' population, the Basic Rorschach Score was not adequate for use in diagnosing individual patients.

Brief Report
Received November 29, 1954.

[1] An extended report of this study may be obtained without charge from Ben C. Finney, VA Hospital, Palo Alto, California, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4481 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

References

1. Buhler, Charlotte, Lefever, D. W., Kallstedt, F. E., & Peak, H. M. *Development of the Basic Rorschach Score: supplementary monograph.* Los Angeles, Calif.: Rorschach Standardization Study. No. 1, 1948.
2. Cass, W. A., & McReynolds, P. A contribution to Rorschach norms. *J. consult. Psychol.*, 1951, 15, 178–184.
3. McReynolds, P. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.*, 1951, 46, 131–141.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Psychotherapeutic Technique, Professional Affiliation, and Experience Level[1]

Hans H. Strupp[2]
Washington, D. C.

To the extent that psychotherapy aspires to scientific status, it must be based upon rational procedures that are capable of precise formulation and communication. A corollary of this proposition is that psychotherapeutic operations are not teachable unless the requirement of communicability is met. To this end, psychotherapy must strive to make its working assumptions, hypotheses, and operations increasingly explicit and objective.

Previous research in this field has dealt primarily with changes in the client as a result of therapy, diagnostic studies of the person seeking help, and the development of process measures. Few investigations have concerned themselves with the person of the therapist, and the character of his contribution to the dyadic interpersonal situation known as psychotherapy.[3] Yet it is almost axiomatic—and has been recognized in psychoanalytic writings since the time of Freud—that the therapist's personality, attitudes, past interpersonal experiences, and emotional blind spots are among the prime determiners of his therapeutic operations. It is believed that a program of research focusing on the therapist's contribution to the treatment situation is potentially productive for advancing our knowledge of psychotherapy.

As a first step, it is proposed to conduct a series of exploratory studies directed at the therapist's verbal techniques. This research attempts to answer the questions: (*a*) What techniques are used by psychotherapists of two theoretical orientations (Rogerian and psychoanalytic)? (*b*) Are there systematic differences in technique attributable to professional affiliation and level of experience? (*c*) Are there similar differences with respect to the therapeutic problem to which the therapist is addressing himself at the moment? (*d*) What is the effect of the therapist's personal analysis upon his verbal operations? The first problem has been dealt with in the first article of this series (6); the second will occupy our attention in the present paper; the remaining ones will be the subject of subsequent publications.

Method

It was considered advisable to avoid the complexities of the actual therapeutic situation, and to study therapists' verbal behavior through the use of a preselected sample of patient communications. Thus, a series of 27 short paragraphs of patient statements was culled from published therapeutic interviews, and typed on individual cards. A bare minimum of background information preceded each statement. Included in the series, which was considered a fair cross section of verbalizations heard from neurotic patients in early interviews, were a variety of complaints,

[1] This article is the second in a series of investigations concerned with the objective study of psychotherapists' verbal operations, and represents one phase of the author's doctoral dissertation submitted to the Graduate Council of the George Washington University (5). Grateful acknowledgment is made to Professor Curtis E. Tuthill, Dr. E. Lakin Phillips, and Leon Salzman, M.D., for valuable suggestions and criticisms.

[2] While carrying out this work, the author held positions successively with the Department of the Air Force and the Department of the Army. Neither organization has any connection with this research.

[3] In Mowrer's (3) recent compendium only one chapter (Fiedler's) specifically addresses itself to this problem.

Table 1

Scoring Categories and Response Frequencies

Category no.	Interaction process analysis definition	"Psychotherapeutic" definition*	Psychiatrists		Psychologists		Social workers	
			n	%	*n*	%	*n*	%
1	*Shows solidarity*, raises other's status, gives help, reward.	Gives reassurance, encouragement, shows compassion, tenderness.	41	4.0	13	5.3	35	10.1
2	*Shows tension release*, jokes, laughs, shows satisfaction.	Not applicable.†						
3	*Agrees*, shows passive acceptance, understands, concurs, complies.	Shows passive acceptance, understanding, is permissive (includes silences).	184	18.1	39	15.8	70	20.3
4	*Gives suggestion*, direction, implying autonomy for other.	Proposes course of action, defines ("structures") the therapeutic situation.	65	6.4	15	6.1	16	4.6
5	*Gives opinion*, evaluation, analysis, expresses feeling, wish.	Interprets, analyzes behavior patterns; thought in process, inferential reasoning.	164	16.1	25	10.0	23	6.7
6	*Gives orientation*, information, repeats, clarifies, confirms.	Restates, clarifies, "reflects" (Rogerian response).	108	10.6	35	14.2	59	17.1
7	*Asks for orientation*, information, repetition, confirmation.	Asks factual questions, expresses lack of knowledge, uncertainty.	31	3.0	17	6.9	18	5.2
8	*Asks for opinion*, evaluation, analysis, expression of feeling.	Explores, asks for elaboration or expression of feeling (includes open-ended questions, "nondirective leads").	367	36.2	88	35.6	118	34.3
9	*Asks for suggestion*, direction, possible ways of action.	Not applicable.†						
10	*Disagrees*, shows passive rejection, formality, withholds help.	Shows passive rejection, disbelief; ignores requests or complaints; thwarts, frustrates.	44	4.3	13	5.3	5	1.4
11	*Shows tension*, asks for help, withdraws out of field.	Not applicable.†						
12	*Shows antagonism*, deflates other's status, defends or asserts self.	Shows antagonism, aggression, sarcasm, irony, cynicism.	13	1.3	2	.8	1	.3
	Total		1,017		247		345	

* This column gives examples of the kinds of responses by therapists included in the categories, which were in no way redefined.
† The absence of responses in these categories is probably a function of the experimental conditions under which responses were elicited.

statements by a seriously disturbed near-psychotic patient, suicide threats, and other transference reactions, such as blocking, negativism, requests for direct advice, and open hostility.

Procedure and Subjects

Therapists who had agreed to participate were visited and interviewed individually. Each interview lasted about one and one-half to two hours.

The respondent was presented with the experimental series of patient statements and requested to state what response, if any, he would make to the hypothetical patient. The therapist was permitted to give as many or as few responses as he chose; "silent" responses were also considered acceptable.[4] Therapists

[4] In contrast to the related study by Phillips and Agnew (4) which restricted responses by a multiple-choice answer form.

were asked to assume that the patient statements occurred early in treatment, that the problem came up for the first time, and that none of the patients were hospitalized. All responses and comments were recorded by the investigator.

Information concerning the therapist's training, experience, theoretical orientation, etc. was collected at the end of the interview.

The subjects were 25 psychiatrists, 7 psychologists, and 9 psychiatric social workers. All claimed allegiance to psychoanalytic or neo-Freudian principles, and most of them had received training from, or reported they had been influenced by, the Washington School of Psychiatry—a powerful agent molding the "therapeutic climate" in the Washington, D. C. area.

Fifteen psychiatrists, 3 psychologists, and 7 social workers had five or more years of experience in psychotherapy and will be treated as "experienced therapists" in subsequent analyses. While such a dichotomy implies a certain arbitrariness, the distinction is quite clear cut in the case of psychiatrists: the experienced therapists were for the most part fully qualified analysts engaged in private practice of psychoanalysis, whereas the inexperienced group ($N = 10$) was made up largely of psychiatric residents with relatively little experience in therapy. The former group thus includes the greatest concentration of professional experience.

Psychologists and social workers were typically employed by local mental hygiene clinics, and in addition to psychotherapy performed diagnostic and related functions.[5]

Results

Quantification of therapists' responses was accomplished by means of Bales' (1) system of interaction process analysis, a well-known general purpose framework for describing social interactions. The system commended itself for present purposes because of (*a*) the painstaking labor that went into its development, and (*b*) its theoretical neutrality with respect to the teachings of conflicting "schools" of psychotherapy.

[5] A systematic comparison between the techniques of these psychoanalytically oriented psychologists and a comparable group of Rogerians has been presented previously (6).

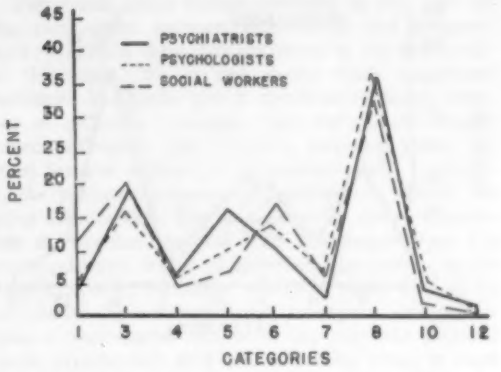


Fig. 1. Response distributions of psychotherapists by professional affiliation.

Table 1 presents Bales' 12 categories, together with the response frequencies of psychiatrists, psychologists, and social workers.

The reliability of the writer's categorizations was tested upon a stratified random sample of 10 cases comprising a total of 370 score units. Two judges who had been trained in the Bales method independently scored the ten "test" cases. Average rater agreement was 78 per cent.

Figure 1 presents the frequencies shown in Table 1 in profile form. Table 2 attempts a systematic comparison between the professional groups with respect to the percentages in any one response category. The *t* values (computed by the formula for uncorrelated

Table 2

Comparisons (*t* Values) Between Professional Groups

Response category	Psychiatrists vs. psychologists		Psychiatrists vs. social workers		Psychologists vs. social workers	
1	.93		4.36**		2.09*	
3	.85	*P*	.92		1.41	
4	.18	*P*	1.20	*P*	.83	*Ps*
5	2.40*	*P*	4.48**	*P*	1.48	*Ps*
6	1.64		3.25**		.97	
7	2.79**		1.83		.85	*Ps*
8	.15	*P*	.63	*P*	.35	*Ps*
10	.67		2.42*	*P*	2.79**	*Ps*
12	.63	*P*	1.50	*P*	.83	*Ps*

P—The response frequency of psychiatrists exceeds that of the other group.
Ps—The response frequency of psychologists exceeds that of the other group.
* Significant at the .05 level using a two-tailed test.
** Significant at the .01 level using a two-tailed test.

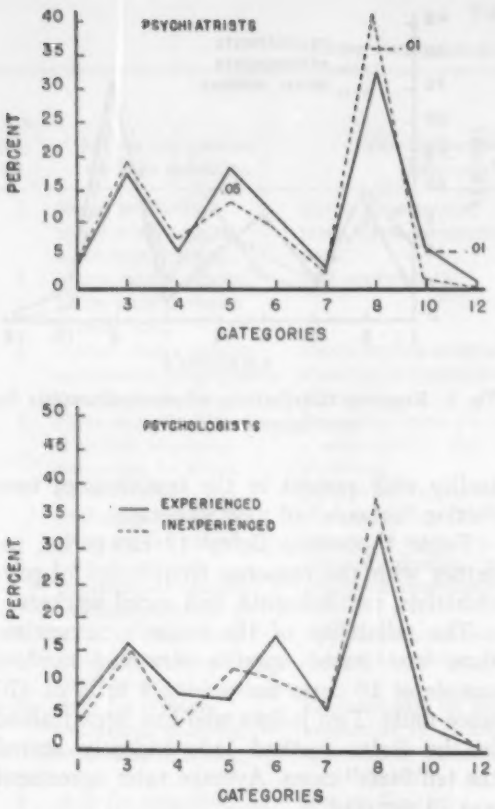


Fig. 2. Response distributions of experienced and inexperienced psychotherapists. (Experienced psychiatrists $N = 15$, $n = 595$; inexperienced psychiatrists $N = 10$, $n = 422$; experienced psychologists $N = 3$, $n = 117$; inexperienced psychologists $N = 4$, $n = 130$.)

percentages) disclose the following statistically significant differences:

1. Psychiatrists tend to give a larger number of interpretive responses (Category 5) than either psychologists or social workers.

2. Psychiatrists and psychologists exceed psychiatric social workers in passive rejections (Category 10).

3. Psychiatric social workers make greater use of reassurance (Category 1) than either psychiatrists or psychologists.

Inconsistent differences in Categories 6 and 7, while possibly quite real, have been disregarded for the present.

The next analysis addresses itself to the question: Are there significant differences between the response distributions of experienced and inexperienced therapists within a professional group? In other words: What is the effect of length of professional experience upon technique?

Figure 2 shows the results in profile form, significant differences being indicated in the graph by the appropriate significance level. Social workers, while showing a response pattern similar to psychologists, have been omitted because of the very small N for inexperienced practitioners.

The distributions are again rather similar; however:

1. Experienced psychiatrists use more interpretations and a larger number of passive rejections.

2. Inexperienced psychiatrists reveal a predilection for exploratory responses (Category 8).

With respect to psychologists, none of the differences are statistically significant, but the aforementioned shift in Category 8 seems to be in the same direction.

Discussion

The most striking finding emerging from the comparison between psychotherapists of different professional affiliations but comparable theoretical orientation is the large degree of similarity of their response profiles.

Of the three statistically significant differences between professional groups, only one (reassurance) appears to be attributable to professional affiliation. The second (interpretation) is more adequately accounted for on the basis of experience level, and the third (passive rejection) is somewhat in doubt.

1. One may regard the preponderance of reassuring responses by social workers as a move away from insight-producing "uncovering" therapy in that this technique assuages but actually changes little. This is not to say that a certain amount of reassurance may not be desirable, beneficial, or useful in the early stages of therapy, particularly with a seriously disturbed or very anxious patient.

2. The finding that psychiatrists exceed psychologists and social workers in the number of passive rejection responses is a provocative one, but because of the relatively small frequencies it is merely reported as suggestive of a possibly important difference (see below).

With regard to experience level, it is interesting to note that psychiatrists show somewhat greater divergence than psychologists. This may be a function of the greater differential in experience between experienced analysts and psychiatric residents than between psychologists of varying degrees of professional experience. While the sample of psychologists was small, the group appears to be fairly homogeneous as long as we disregard Rogerian therapists whose techniques, as we have shown elsewhere (6), are radically different.

The principal technique differences between experienced and inexperienced psychiatrists concern (*a*) interpretations, (*b*) explorations, and (*c*) passive rejections.

a. As a psychiatrist becomes more experienced, he appears to place greater emphasis upon interpretations. It was also demonstrated (5) that this characteristic not only differentiates reliably experienced psychiatrists from inexperienced ones but also from psychologists and social workers.

The significance of this result is twofold: (i) it attests to the emphasis given to interpretations by experienced psychiatrists in that this professional group follows, perhaps more closely than any other, psychoanalytic doctrine, which extols interpretation as the hallmark of intensive psychotherapy; (ii) it contravenes to some extent the recommendations of such authorities as Fromm-Reichmann (2) who caution against interpretations before sufficient information to document this therapeutic maneuver has been obtained. Thus, psychiatrists, as the most thoroughly trained group of therapists, might be expected to be more wary of interpretations, but if such cautiousness does exist, it is not revealed by the present data.

b. Concerning the statistically highly significant difference between experienced and inexperienced psychiatrists in Category 8, the following observations appear warranted. Inexperienced therapists display a tendency to ask more exploratory questions, which by and large are made at the expense of interpretations and passive rejections. It appears altogether reasonable for a therapist when confronted with a patient about whom he knows little to ask for further information; but here we have objective evidence that inexperienced psychiatrists reveal a more pronounced tendency to develop a patient's theme along exploratory lines whereas experienced psychiatrists proceed more readily to interpretations. While we cannot answer the question as to which is more effective therapeutically, there is at least the suggestion that to the extent that explorations are a sign of inexperience, less experienced therapists with medical training subscribe to the precept: If in doubt, ask for more information.

This result gains further meaning in the light of the comparison between experienced and inexperienced Rogerian therapists presented in the first study of this series (6). It was shown there that inexperienced Rogerians give a significantly larger number of reflective responses than their more experienced colleagues, the reflective response being the most favored variety. In the present study, Category 8, the preferred response of psychiatrists, reveals the same relationship. The implication is clear: Regardless of whether we deal with the Rogerian or the psychoanalytic frame of reference, the most popular (stereotyped?) technique derives a good deal of its popularity from inexperienced therapists, and undergoes a pronounced decline as the therapist becomes more experienced. It is true that this trend is more pronounced for Rogerian psychologists than for analytically oriented psychiatrists, but the same conclusion seems justified: An increase in experience leads to a diversification of therapeutic technique, and, conversely, reliance on one specific technique is a characteristic of inexperience.

c. Because of the small frequencies, it is difficult to say whether passive rejection is a technique preferred by experienced psychiatrists or by psychiatrists in general, although we are inclined to accept tentatively the former interpretation.

Summary and Conclusions

This is the second in a series of investigations to elucidate the psychotherapist's contribution to the treatment situation. It attempts a systematic comparison between the techniques used in early interviews by psychiatrists, psychologists, and psychiatric social workers subscribing to psychoanalytic principles. Therapeutic responses were secured from 25 psychiatrists, 7 psychologists, and 9 psychiatric social workers of varying degrees of professional experience by presenting a series of 27 patient statements extracted from actual therapeutic interviews. A total of 1,609 response units were categorized by Bales' system of interaction process analysis. Average rater agreement was 78 per cent. Salient results may be summarized as follows:

1. The response profiles of the three professional groups manifest a considerable degree of similarity. Of the statistically significant differences, only one appears to be a function of professional affiliation, namely the psychiatric social workers' predilection for reassurance.

2. Comparisons between experienced and inexperienced practitioners among psychiatrists and psychologists (social workers were omitted because of a small *N* for inexperi-

enced workers) disclose the following reliable differences: (*a*) experienced psychiatrists give a larger number of interpretive responses; (*b*) inexperienced psychiatrists show a preference for exploratory responses; (*c*) experienced psychiatrists use more passive rejections.

It is concluded that professional affiliation exerts a relatively minor influence upon the kinds of techniques used by psychiatrists, psychologists, and psychiatric social workers. As long as the variable of theoretical orientation is held constant, all therapists adhering to psychoanalytic principles employ very similar techniques. When experience level is taken into account, psychologists emerge as a somewhat more homogeneous group in contradistinction to psychiatrists for whom more conspicuous differences between experienced and inexperienced therapists are in evidence. Intensive training in psychotherapy seems to stress interpretations at the expense of exploratory (probing) responses and leads to a general diversification of technique. Previous evidence from Rogerian therapists provides added support for this latter conclusion.

Received July 10, 1954.

References

1. Bales, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley Press, 1950.
2. Fromm-Reichmann, Frieda. *Principles of intensive psychotherapy.* Chicago: Univer. of Chicago Press, 1950.
3. Mowrer, O. H., *et al. Psychotherapy: theory and research.* New York: Ronald, 1953.
4. Phillips, E. L., & Agnew, J. W. A study of Rogers' "reflection" hypothesis. *J. clin. Psychol.,* 1953, **9**, 281–284.
5. Strupp, H. H. An empirical study of certain psychotherapeutic operations: an exploration of the verbal response techniques of psychiatrists, psychologists, and psychiatric social workers. Unpublished doctor's dissertation, George Washington Univer., 1954.
6. Strupp, H. H. An objective comparison of Rogerian and psychoanalytic techniques. *J. consult. Psychol.,* 1955, **19**, 1–7.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Adjustment Changes in the Absence of Psychotherapy[1]

Rosalind F. Dymond
Counseling Center, University of Chicago

It has often been argued that there is no evidence apart from the feelings and impressions of psychotherapists themselves that psychotherapy produces changes that are any "greater" or "different" from those which would have resulted in about the same proportion of cases if they had not been seen by a psychotherapist. It is common knowledge that many people in severe psychological distress "pull out of it" or "spontaneously remit" without having been so treated.

Procedure

In the research program evaluating client-centered psychotherapy (6) some attempt was made to study this phenomenon. Nineteen of the clients who applied for psychotherapy at the Counseling Center of the University of Chicago were required to wait before beginning their interviews. These subjects were given a battery of psychological tests when they first applied and were retested with the same battery after a two-month waiting period. Admittedly two months is much too short a time to really test the possibility of spontaneous remission. It was selected because this was the average period that the non-research cases had to wait for a therapist. It was felt that it would be unfair to people in trouble to put off their therapy longer than this for research purposes when they felt the need for immediate help.

After waiting for the two-month period, 11 of these people went on to complete their psychotherapy (six or more interviews), and were tested again with the same battery for the third time. The remainder of the group, eight people, came a time or two and then failed to return. These we came to call the "attriters."

Grummon (5) has reported that there are some very interesting differences between the tests done before and after the waiting period for these two groups. He found that those who waited and then completed therapy showed no improvement over the waiting period on a variety of tests, but those who waited and then failed to get started in therapy before breaking off showed that significant improvement had taken place during their waiting period.

This finding raises a number of questions:

1. In what respects did the attrition group change and the therapy group fail to change?
2. Were the two groups drawn from the same population or were they differentiable to start with?
3. Were the changes in the attrition group of the same kind as those which are characteristic of successful psychotherapy subjects?

Results

Questions 1 and 2 will be taken up first, and the remainder of this paper will be concerned with question 3.

Table 1 shows that in terms of formal characteristics the attriters do not appear to be markedly different from the therapy subjects in terms of age, sex, and occupation (student or nonstudent).

The tests on which the attrition group showed the most marked improvement were those which involved self-description. The most dramatic of these was the *Q*-adjustment score (2). This was the scoring for adjustment of each subject's sorting of 100 self-referent items along a continuum of "this statement is very like me" to "this statement

[1] This investigation was supported (in part) by a research grant (PHS M903) from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service.

Table 1

Wait Group Subjects

Group	Age		Sex		Occupation*		Number of interviews	
	Mean	Range	M	F	S	NS	Mean	Range
Therapy ($N=11$)	27.4	22–36	7	4	8	3	36.6	6–108
Attrition ($N=8$)	24.6	20–29	3	5	3	5	2.1	1–5

*S = student; NS = nonstudent.

is very unlike me." The basic instrument was developed by Butler and Haigh (1). Table 2 gives the prewait and pretherapy scores of the two groups.

The difference between the mean scores of the therapy group at the two testing points was not significantly different using t for correlated means, but for those who did not go into therapy the difference between the prewait and pretherapy means was significant at the 2% level, using the same statistic. The difference between the means of the two groups at the first testing was not significant. These data seem to support the conclusion that, although the groups were about equally maladjusted initially as self-described, eight of them managed to improve while waiting for therapy and then chose not to continue their interviews. The 11 who did not improve while waiting for therapy to begin, as measured from their self-description, did go on to complete therapy. However, all these conclusions are based on what the individual is able or willing to say about himself.

Another instrument which formed part of the battery administered before and after the two-month wait was a standard 20-card Thematic Apperception Test (TAT). The

Table 2

Q-Adjustment Score of Therapy and Attrition Groups

Group	Prewait		Pretherapy	
	Mean	Range	Mean	Range
Therapy	28.6	16–38	30.2	15–41
Attrition	32.0	22–46	42.9	25–55

prewait and pretherapy records of the 19 subjects were mixed in with the records of all the other subjects in this experimental program (6). They were coded and all identifying material was removed. A mental health rating was then made for each record as a whole (4). Each record was rated from 1 to 7 with a rating of 1 being equivalent to "bordering on psychotic or psychotic," and 7 being a "well-integrated, happy person, socially effective." Table 3 compares the two wait groups in terms of their TAT ratings.

Here, none of the t's reached significance, showing that the groups were equally maladjusted at the initial testing point and neither improved significantly over the waiting period. This result is sharply at variance with the results of the *Q*-adjustment measure. What is the relation of these two adjustment scores? Rho correlations were run between

Table 3

TAT Mental Health Ratings of Therapy and Attrition Groups

Group	Prewait		Pretherapy	
	Mean	Range	Mean	Range
Therapy	3.0	1–4	3.4	1–5
Attrition	3.6	3–5	3.7	3–5

the TAT ratings and the *Q*-adjustment scores for both groups at the two testing times. The rho's for the therapy group were + .47 at prewait and + .30 at pretherapy. For the attrition group, the rho at prewait was + .39 but at pretherapy it was − .20! These results and those of a previous study (3) would seem to show that although there is ordinarily a fair relationship between these two measures of adjustment, this relationship does not hold for the attrition group at pretherapy time. This seems to cast some doubt on the apparent improvement in the waiting period of these people.

The lack of support from the TAT analysis for the positive adjustment change in self-description raises the question: Are the self-descriptive changes defensive in nature, perhaps in response to the threat of therapy? The question brings us to question 3: Of what does the improvement in the absence of

Table 4

Comparison of Attrition and Success Groups on *Q* Adjustment and TAT Ratings

Group	*Q* adjustment			TAT rating		
	First score	Second score	Change	First score	Second score	Change
Attrition ($N = 6$)						
Mean	33.0	47.3	+14.37**	3.5	3.5	0
Range	22–46	40–55		3–4	3–5	
Success ($N = 6$)						
Mean	34.83	48.33	+13.50**	3.8	5.0	+1.2*
Range	23–42	35–54		3–5	4–6	

Note.—For the attrition group the first score was their score at prewait and improvement was change from prewait to pretherapy, while the success cases' first score was established at pretherapy and improvement was change from pretherapy to posttherapy.

* Significant at 5% level.

** Significant at better than 1% level.

therapy consist? Is the improvement in the attrition group during the no-therapy period of the same kind as that which results from a period of successful psychotherapy? The pattern of large *Q*-adjustment–score increase unsupported by commensurate change in the TAT rating did not hold for all eight of the attriters to an equal extent. However, for six of them the pattern was clear. These six were matched with a group of successful therapy cases in terms of their initial status and degree of improvement on the *Q*-adjustment score. "Success" was established by a threefold criterion: (*a*) *Q*-adjustment–score increase from pretherapy to posttherapy, (*b*) TAT-rating increase from pretherapy to posttherapy, and (*c*) the therapist's rating of the case on a nine-point scale of improvement at six or above (7). The matching data are shown in Table 4.

To make a start on the problem as stated in question 3 above, the original self-descriptive *Q* sorts were examined. Each of the subjects had sorted the 100 self-referent items at the different testing points into nine piles under the instructions, "Sort the cards to describe yourself as you see yourself today, putting the card that is least like you at the 'zero' end and the one that is most like you at the 'eight' end, then the four cards that are next most and least like you at 'one' and 'seven' and so on." The required distribution is shown in Table 5.

Table 5

Q-Sort Distribution

	Least like me								Most like me
Pile number	0	1	2	3	4	5	6	7	8
No. of cards	1	4	11	21	26	21	11	4	1

Each item was given the weight of the pile number into which it fell in the particular sorting. The average sorting that the six attrition subjects made at the pretherapy point was then computed and compared to the average sorting made by the six matched success cases at their posttherapy point. The differences between the two groups in the mean weighting of each item were tested for significance using *t*. Table 6 lists the items which were significantly different with respect to their average weighting.

Although the smallness of the *N*'s on which these results were based puts some stricture on the degree of interpretation they warrant, the picture of the attrition group that emerges from a reading of the items is so clear that it seems probable that something like it would be found in a study with more adequate samples. The attriters appear to be a group of "rugged individualists" who are hanging on tight to the idea, "I *am* someone, I *am* superior, I *do* want to cope and will do it *myself*." They seem to have difficulty accepting the idea of help from others (I am an aloof, reserved person; I make strong demands on myself; I am self-reliant, competitive; I have

Table 6

Q Items Which Differentiate Attrition and Success Groups

Item	Mean weighting: Success	Mean weighting: Attrition	*p*
Weighted *more* characteristic by attrition group			
I am an aloof, reserved person.	2.3	4.3	.01
I am ambitious.	4.3	6.0	.02
I feel superior.	3.5	4.6	.02
I am self-reliant.	4.3	5.5	.05
I am afraid of sex.	2.0	3.5	.05
I just don't respect myself.	2.1	3.1	.05
I am a competitive person.	3.6	4.8	.10
I make strong demands on myself.	3.6	5.0	.10
I don't trust my emotions.	3.0	4.5	.10
I have to protect myself with excuses, with rationalizing.	3.1	4.3	.10
Weighted *less* characteristic by attrition group			
I want to give up trying to cope with the world.	3.5	1.1	.01
I am no one. Nothing really seems to be me.	3.8	1.6	.02
I am liked by most people who know me.	6.8	5.6	.05
I understand myself.	6.0	4.3	.05
I express my emotions freely.	5.6	4.0	.10

to protect myself with excuses, with rationalizing). It does not seem surprising in the light of this pattern that they have made a strong effort during the waiting period to pull themselves up by their own bootstraps; they have decided that therapy is not for them and do not follow it through; and their TAT's do not show any deep personality change.

What are the characteristics of the successful psychotherapy subjects in contradistinction to the attriters who "go it alone"? Some answer is given by the items: I express my emotions freely, I am liked by most people who know me, I trust my emotions, I respect myself, I understand myself. All of these were more characteristic of the self-description of the successes than of the attriters.

Summary

This study has been a bare beginning on a question which has been an open one far too long. Although there are many shortcomings in the study presented here, some conclusions can be drawn. A summary may help to clarify the points to be made.

1. There was a small group of six subjects who appeared to have "spontaneously recovered," that is, decided after waiting for therapy that they did not want therapy after all, and appeared to have improved in adjustment, as measured from their self-description, about as much as those who went through therapy successfully. This group was not initially differentiable in terms of adjustment status from the subjects who did *not* improve in adjustment while waiting and who *did* go into therapy.

2. The attriters' increase in adjustment as measured from the self-descriptive *Q* sorts was not confirmed by a commensurate increase in mental health as rated blindly from their TAT's. On the basis of the TAT ratings, "the spontaneous remission" group appeared to be as poorly adjusted at the end of the waiting period as they were when they first applied for therapy, not significantly different in adjustment from those who then continued with therapy, and significantly poorer in adjustment than those who completed therapy. In other words, in terms of the adjustment scoring of the subject's own self-description the attriters had improved, whereas in terms of the scoring of their projective tests by a diagnostician they had not.

3. To discover whether the "improved" self-description of the attriters was differentiable from the improved self-description of a group which had completed therapy successfully, the attriters were matched with a successful therapy group in terms of initial status and degree of improvement on the *Q*-adjustment measure. From a comparison of the average "improved" self-description of these two groups, 15 items were found which were weighted in terms of "like me" or "unlike me" significantly differently. A study of these crucial items appeared to show that the "improved" self-description of those who did not go into therapy had strong defensive and compulsive elements which were absent or significantly less characteristic of the self in successful cases as described by them.

This study, limited as it is by the small number of cases and short period of time covered, has shown that although positive adjustment changes appear to take place in maladjusted persons in the absence of psychotherapy, these are not identical to the changes which occur in equally maladjusted persons who complete therapy. No "deep" reorganization appears to take place. The "improvement" appears to be characterized by a strengthening of neurotic defenses and a denial of the need for help.

In future studies on this subject, not only should the sample be enlarged and the period of no therapy be extended, but more data should be gathered: additional diagnostic data to discover the nature of the defense structure of those who do and do not continue with therapy, and follow-up data so that the effects of self-improvement of this kind may be compared to the effects of successful psychotherapy.

Received July 16, 1954.

References

1. Butler, J. M., & Haigh, G. V. Changes in the relation between self-concepts and ideal concepts. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 55–75.
2. Dymond, Rosalind F. An adjustment score for *Q* sorts. *J. consult. Psychol.*, 1953, **17**, 339–342.
3. Dymond, Rosalind F. Adjustment changes over therapy from self-sorts. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 76–84.
4. Dymond, Rosalind F. Adjustment changes over therapy from Thematic Apperception Test ratings. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 109–120.
5. Grummon, D. L. Personality changes as a function of time in persons motivated for therapy. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 238–255.
6. Rogers, C. R., & Dymond, Rosalind F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954.
7. Seeman, J. Counselor judgments of therapeutic process and outcome. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 99–108.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Language and the Interpretation of "Sexual" Symbolism[1]

Harold H. Mosak
Roosevelt University

In current psychoanalytic practice the interpretation of symbols is accorded two types of treatment. Some psychoanalysts believe that certain symbols are fixed symbols while others regard these symbols as meaningful only when interpreted in the context of the individual or group using these symbols. Freud advocated the first position in his statement that "there are symbols which preponderantly, or almost exclusively, designate one of the sexes, and there are yet others which . . . have only the male or only the female signification. To use long, stiff objects and weapons as symbols of the female genitals, or hollow objects . . . as symbols of the male genitals is not permitted by the imagination."

Freud's statement would lead us to expect that language, also a symbolic form, should bear the stamp of these unconscious processes. Consequently, a list of "sexual" symbols was translated into three languages—French, Spanish, and German—which designate gender for nouns. If Freud's statement is valid, we might assume that symbol and gender designations would coincide. Our results indicate that Freud's contention is not borne out in language. In French 15 of the 33 "male" symbols possess feminine gender while 12 of the 26 "female" symbols are designated by masculine articles. Spanish assigns feminine articles to 19 male symbols and masculine articles to 12 female symbols. In German symbol-gender reversals occur for 11 of the male symbols and 9 of the female symbols. We also note that the designation of "long, stiff objects and weapons" as feminine and "hollow objects" as masculine is indeed "permitted by the imagination." No consistency in gender appears for the three languages. Only 21 of the 59 symbols are consistently used with the same gender, and of these 21, 8 demonstrate consistent symbol-gender reversals. These results warrant seriously questioning whether these symbols are either fixed or sexual. It might be argued that Freud's theory of symbols was based on his expressed feeling that dreams and dream elements are preponderantly visual images and therefore would not possess parallelism with language symbolism. Freud himself confutes such an argument when he uses etymological evidence to corroborate his theory of dream symbolism.

In the area of psychotherapeutic investigation such results as ours confront us with an anomalous situation. For despite lack of theoretical foundation or experimental verification, therapeutic techniques may nevertheless prove effective. This effectiveness might even be utilized to demonstrate the *validity* of the assumptions underlying the use of the technique. Actually, this effectiveness merely demonstrates the *utility* of the technique. In the framework of psychoanalytic theory, Freud's hypothesis might possess such utility. For other schools of psychotherapy the theory of fixed sexual symbols would presumably possess little utility or validity.

[1] An extended report of this study may be obtained without charge from Harold Mosak, 4305 W. Flournoy St., Chicago, Ill., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4466 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Brief Report
Received November 22, 1954.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Conformance in Role Perception as a Personality Variable[1]

Theodore R. Sarbin and Curtis D. Hardyck

University of California, Berkeley

This study, one of a series growing out of a social psychological theory of behavior (5), attempts to clarify "conformance," a perceptual variable, and throw some light on its personological correlates.

Conformance is defined as the modal perceptual response to a social stimulus object. It is parallel to the notion of conformity as applied to overt conduct. When a social object elicits qualitative perceptual responses from members of a specified subculture, a relatively large proportion of the reference population will agree on the quality represented by the object; a small number will differ from the majority but will agree among themselves; and a few will agree with virtually no one. Figure 1 illustrates the notion graphically.

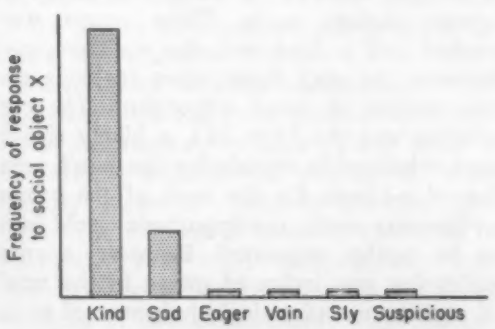


Fig. 1. An illustration of conformance.

The stimulus material employed in this study was the same as that developed for a taxonomy of postural cues used in role perception and inference (6). We prepared a set of 44 schematized "stick figures" of humans in various postures. The figures were drawn by an interpretive dancer and represent a wide range of expressive and attitudinal states. The figures were so constructed as to offer no cues for interpretation other than posture. Figure 2 is one of the items in the Stick Figures Test.

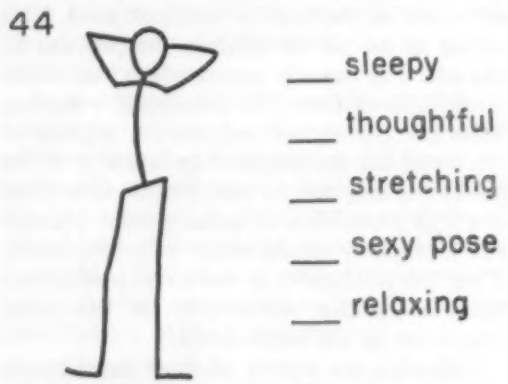


Fig. 2. Illustrative item from the Stick Figures Test.

As a preliminary step the figures were presented on 3 × 5 lantern slides to a group of 150 upper-division psychology students. The students were given the following instructions:

> A number of stick figures, representing people in various postures, will be shown to you. As the figures are shown, you are to write 1 or 2 adjectives which best describe your impression of that figure.

From the data obtained in this sample, a check list of five adjectives, ranging from most frequent to least frequent, was constructed for each figure. As no significant sex differences were found, men and women were combined in preparing the check lists. The reliability of the adjective lists was determined by taking a second sample of 70 upper-division psychology students and repeating the procedure given above. No significant differ-

[1] No. 7 in a series entitled "Contributions to Role-taking Theory." The authors wish to express their gratitude to Professor H. G. Gough, Drs. Frank Barron, Albert Kostlan for serving as judges.

ences in proportion or kind of adjectives were found.

In developing norms for the use of the stick figures, we noted impressionistically that certain persons characteristically selected nonmodal, i.e., nonconformant responses. From these impressions and from certain implied postulates of role theory, we were led to the following hypothesis: persons who frequently make nonconformant perceptual responses are characterized by social behavior which is deviant or socially invalid.

The theoretical basis for this hypothesis follows from notions originally advanced by G. H. Mead (4) and more recently developed by the senior author (5). Role perception is the first part of a social act; it is the (usually) silent naming or locating of the position of the other on the basis of available cues. This serves as an aid in locating the position of the self. The motoric response, the role enactment, follows from the perceptual response. Since the perceptual and motoric aspects of the social act are acquired holistically, if the perceptual response is conformant, then there is a high probability of socially valid (though not necessarily conforming) role enactment. Thus, the veridicality of one's role perceptions bears a positive relationship to functional adaptation in the social world.

Following are reports of three experiments which were carried out to test the hypothesis.

Schizophrenic vs. Normal Responses

As an initial check on the validity of our procedure, the stick figures were administered individually to 30 cooperative, undeteriorated, hospitalized schizophrenics. They were given the following instructions:

> A number of stick figures, representing people in various postures, will be shown to you. Each figure will be numbered, corresponding to a number on this paper. For the corresponding number on this paper, there is a list of five words, plus a sixth space which is left blank. As the figures are shown to you, you are to check the word which you think best describes your impression of that figure.
>
> If the feel that none of the words on the list adequately describes your impression of the figure, write what you consider to be the best description in the space marked "other."

No time limit was set, and the experimenter was allowed to rephrase the instructions if a subject failed to comprehend the initial instructions.

The distributions of responses for the schizophrenic sample were compared with the norms obtained on our previous samples. For the schizophrenic sample the distributions were all uniformly rectangular, in comparison to the J curves obtained from college normals, thus offering substantial evidence for the usefulness of the technique and for the initial confirmation of the hypothesis.

Conformance Scores and Social Adjustment

In the second exploratory study, because of practical limitations, we were unable to obtain direct behavioral rating such as would be obtained in an assessment program. Instead we made use of indirect sources of information on role enactment. The California Psychological Inventory (Gough, 2) has a number of scales, all of which are empirically rooted in social behavior. This instrument was administered to 21 undergraduates who had been subjects for the Stick Figures Test. The Inventory profiles were then ranked by Gough according to a "social adjustment" criterion.

Responses to the stick figures were scored as follows: The modal response to each stick figure for the previous samples was assigned a weight of 1. The subjects were then given a score in terms of the number of modal response choices made. These scores were ranked and a Spearman rho was computed between the stick figure score rankings and the ranking on social adjustment. The correlation was .81 ($p = .01$), a highly significant relationship considering the small number of subjects. On the basis of the second preliminary study, our hypothesis would seem to be further supported. However, a cross validation was indicated owing to the small N and the use of a single judgment of social behavior.

Cross Validation of Conformance and Social Adjustment

The next step involved the presentation of the figures to a sample of 91 upper-division psychology students. The instructions and scoring method were the same as in the two previous studies. For this group the Minnesota Multiphasic Personality Inventory (MMPI)

was used as a measure of personality functioning. The MMPI profiles were sorted by each of three judges into a normal distribution on the following criterion of *Adjustment:*

> Getting along well in the world as it is. Adequate degree of social conformity; capacity to adapt to a wide range of conditions. Ability to fit in (3).

The results were as follows: The range of scores on the Stick Figures Test was from 11 to 26 (44 stick figures). The correlations of the conformance scores on the Stick Figures Test with the sorts of the MMPI profiles were .32 for Judge A, .68 for Judge B, and .52 for Judge C.

Combining of the three sorts to obtain a more stable criterion of social adjustment resulted in a Pearsonian product-moment correlation of .51 with the conformance score. The correlation is unusually high, especially in the absence of a refined method for obtaining the conformance score.[2] The above sorts are done with sex groups combined. Separate sorts for males and females revealed no significant differences among the correlation coefficients.

Interjudge reliabilities on social adjustment were .66 between Judges A and B, .73 between A and C, and .83 between B and C. They are considered adequate for our experimental purposes. The split-half reliability for the Stick Figures Test is .50. Refinement of the test by item analysis will doubtless increase the reliability.

Correlations between conformance score and certain clinical scales provide some additional confirmatory evidence. The *F* scale on the MMPI indicates the number of "infrequently answered" items in the test. In a sense it is a subtle measure of conformance. Pearsonian r is $-.32$. Subjects who score high on the *F* scale tend to preceive the postures in a nonconformant way. Another index of social maladjustment is the number of clinical scales on the MMPI on which the subject's score is two sigmas or more above the mean. This index correlates with conformance $-.33$ (and was probably used implicitly by all the judges in making their judgments of social adjustment). An empirically devised scale of ego strength (1) correlates .50 with conformance.[3]

[2] In view of the method used to obtain the conformance score, the possibility of curvilinearity should be taken into account. Scatter plots of the correlations revealed no significant tendency toward curvilinearity. It would seem safe to conclude, therefore, that judgments in terms of conformance do bear a significant relationship to level of adjustment as measured by the MMPI profile sorts.

Summary

On the basis of certain postulates derived from role theory, a significant relationship between nonconformant perceptual responses and deviant social behavior is hypothesized. Three experiments were performed to test this hypothesis. The first experiment compared the distribution of responses of normals and schizophrenics to a set of schematized stick figures of humans in various postures. The normal subjects were found to have J-curve distributions as contrasted with the rectangular distributions of the schizophrenic sample.

In a second experiment, the number of conformant responses of a group of normal subjects was found to bear a significant relationship to their social adjustment as judged by test profiles. A third, cross-validational, experiment indicated the stability of the relationship. Correlations with certain empirically derived personality scales offer further confirmation of the hypothesis.

Received July 12, 1954.

References

1. Barron, F. An ego-strength scale which predicts response to therapy. *J. consult. Psychol.*, 1953, 17, 327–333.
2. Gough, H. G. *The California Psychological Inventory*. Berkeley: Univer. of California Press, 1951.
3. Gough, H. G. Predicting success in graduate training: a progress report. Unpublished manuscript, Univer. of California, Institute of Personality Assessment and Research, 1950.
4. Mead, G. H. *Mind, self and society*. Chicago: Univer. of Chicago Press, 1934.
5. Sarbin, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 223–258.
6. Sarbin, T. R., & Taft, R. *An essay on inference in the psychological sciences*. Berkeley: Garden Library Press, 1952.

[3] Conformance scores and an indirect measure of kinesthetic inference (empathy) are correlated to the extent of .48. This finding is only peripherally related to the hypothesis and is not discussed in the present paper.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Rorschach Experience Balance in Epileptics

Paul W. Pruyser[1] and Angela T. Folsom
The National Veterans Epilepsy Center, Boston

The Rorschach literature contains several observations of abnormalities in the use of the human movement and color determinants by epileptics. Rorschach (9), using mostly deteriorated epileptics, reported an underproduction of *M* in slowly dementing patients and a marked overproduction of *M* in the most demented ones and concluded that the quantity of *M* in an epileptic record is proportional to the degree of dementia. Rorschach described an overemphasis on the color determinants, also increasing with the degree of dementia, and often formed by the use of many pure *C* responses, including the so-called color-naming (*Cn*) response. Since the publication of Rorschach's manual several workers have partially or wholly corroborated his findings, except for his observation that the degree of *M* increases with more pronounced dementia. In intramural as well as extramural groups of epileptics the finding of low *M* is general (1, 4, 6, 7, 10, 12, 13) and several workers have confirmed the notion of a typically epileptic, extratensive experience balance (EB). A disproportionate emphasis on color, usually with predominance of pure *C* and *CF* over *FC* has been found to contribute to the extratensive tendency (1, 4, 10, 12). These indices have also been used to validate the old, widespread notion of the so-called "emotional explosiveness" of epileptics and have even served as an aid in diagnosing epileptics from Rorschach records (5). As a matter of fact, it appears that Rorschach's own use of the findings of the predominance of unformed color in his group of epileptics played a role in defining the diagnostic meaning of the *CF* and *C* responses as indicating emotional explosiveness (9, Tables I–VII and XII). Obviously, this part of the validation of color responses is based upon a circular argument which starts with a clinical notion and interprets the meaning of an abnormal test response directly in terms of that clinical conviction even before it is systematically verified.

Studies of the intellectual functioning of epileptics have shown that samples from different strata of the epileptic population give very different results. For instance, whereas testing of institutionalized patients shows on the whole a subnormal range of IQ's, samples of extramural patients have yielded a mean IQ and a range comparable to those of the normal, nonepileptic population (3, 13). It is reasonable to assume, therefore, that the selection of samples may also affect the Rorschach findings, the more so since it seems now well established that the frequency of *M* responses is positively correlated with IQ level. To what extent the normal correlation between IQ and amount of *M* holds for epileptics is as yet unclear, since accurate data on IQ's are lacking in several of the earlier Rorschach publications. With few exceptions (6, 7, 13), most authors have published on deteriorated and/or institutionalized groups, or on samples with an abnormal proportion of below average IQ's.

Present Study

The present study offers EB data on a group of 136 male epileptic veterans, all extramural patients, who were hospitalized only for diagnostic study and medical treatment of their symptoms. They are probably a representative group of intellectually grossly intact epileptics. They have a mean Wechsler-Bellevue Full Scale IQ of 102.6 with a range from 78 to 129; their mean age is 31 (range 21–52). In the vast majority the epileptic

[1] Now at Topeka State Hospital, Topeka, Kansas.

symptoms made their first appearance after adolescence. The group permits the following subdivisions:

22 with idiopathic grand mal epilepsy
15 with idiopathic psychomotor epilepsy
23 with idiopathic grand mal and psychomotor epilepsy
39 with symptomatic grand mal epilepsy
9 with symptomatic psychomotor epilepsy
28 with symptomatic grand mal and psychomotor epilepsy

In order to assess the generality of the previously reported abnormalities in the EB of epileptics, which is the purpose of this study, we propose:

a. a statistical comparison of our epileptic group with a nonepileptic group of normal intellectual standing;

b. a more detailed analysis of the components of the EB.

Results

The publication of Rorschach norms by Beck *et al.* (2), based upon the Spiegel sample, provides a useful frame of reference for statistical comparison. Table 1 shows the results of a comparison between Beck's group of normals and our group of epileptics for the determinants which constitute the experience balance. In order to cope with the gross deviations from the normal distribution curve which characterize most Rorschach scores the following procedure was adopted: a median score was computed for the two groups combined [2] and this score was taken as the cutting point for the development of a 2×2 table containing for each group the number of cases falling below and above that point. The p values derived from a chi-square test would then indicate to what extent the null hypothesis, in this case whether the two groups could be seen as samples from the same population, was tenable. The table gives the relevant figures and indicates that significant differences exist between the two groups. For both movement and color responses the epileptic group shows a striking underproductivity, which contrasts with previously reported findings of an overuse of color by epileptics as a group.

Table 1

Comparison of EB Determinants in 136 Epileptics and Beck's Normals

Measure		*N* below combined group median	*N* above combined group median	*p*
M				
	Normals	46	111	.001
	Epileptics	93	43	
FC				
	Normals	44	113	.01
	Epileptics	62	74	
CF				
	Normals	55	102	.02–.01
	Epileptics	66	70	
C				
	Normals	100	57	.01
	Epileptics	109	27	
Sum *C*				
	Normals	47	110	.02–.01
	Epileptics	63	73	

Table 2

Frequency of Each Number of *M*, *FC*, *CF*, and *C* Responses in 136 Epileptics

	Response			
Frequency	*M*	*FC*	*CF*	*C*
0	37	62	66	109
1	56	42	43	18
2	15	13	11	2
3	18	10	8	6
4	5	6	4	1
5	3	3	3	0
6	1	0	0	0
7	1	0	0	0
8	0	0	1	0

As an introduction to a further analysis of these data we specify in Table 2 the frequencies for each number of *M*, *FC*, *CF*, and *C* responses for the total epileptic group. Inspection of the figures shows that there is a sharp increase in the zero category if one moves from *M*, through *FC* and *CF*, to *C*. To assess these data properly and to deal with the possible influence of *R*, we employed the measure *M*% as introduced by Rapaport

[2] We express our indebtedness to Drs. Donald W. Fiske and Samuel J. Beck of the University of Chicago. The former was kind enough to furnish the raw scores of the Spiegel sample which enabled the use of distribution-free statistics.

Table 3

M% in 136 Epileptics and Beck's Normals

Frequency (in per cent) in each class	Group			
	Idiopathic ($N=60$)	Symptomatic ($N=76$)	All epileptics ($N=136$)	Normals ($N=157$)
0	32	24	27	15
1–20	5	5	5	2
21–40	20	18	19	18
41–60	12	22	18	23
61–80	15	17	16	24
81–100	17	13	15	18

et al. (8). It is based on the formula

$$\frac{M}{M + Sum\ C} \times 100$$

computed for each record separately, expressing in the range 0–50 a color prevalence, and in the 51–100 range a movement prevalence. Values between 40% and 60% can be taken as without clear-cut prevalence in either direction. Table 3 shows the percentage of cases falling in each of five intervals of an *M*% scale, with the corresponding figures of Beck's normal group for comparison. For the epileptic group as a whole, 51% of all cases seem to show color predominance, against 31% with a movement predominance. The first figure, however, includes cases in which the *M*% was zero as a result of total absence of *M*. In such cases the *M*% is not a fair index and should not be used; when we exclude these cases from the color-predominant group, only 24% of all cases have a clear-cut color prevalence in the range from 1 to 40. Such a low incidence of color prevalence in our group is in sharp contrast to earlier reports in the literature which show a general predominance of color responses for epileptics. Moreover, a further analysis of those cases which have an *M*% from 0 to 40 shows that 53% of them had no *M*, and 40% had only one *M*, whereas only 7% had two or more *M*. In other words, in the majority of cases with a seeming predominance of color the real abnormality lies in a marked underproduction of *M* rather than in any overproduction of *C*, *CF*, or *FC*. A comparison of the distributions of the total epileptic with the normal group by means of chi square (based on the raw data instead of the percentages shown in the table) yields a p between .02 and .05.

A more familiar approach to EB lies in assigning each case to one of the following four categories, as recently more clearly defined by Stotsky (11):

1. Coarcted: if both *M* and Sum *C* are equal to one or less
2. Ambiequal: if either *M* or Sum *C* are greater than one and the difference between them is one or less
3. Introversive: if *M* is greater than Sum *C* plus one
4. Extratensive: if Sum *C* is greater than *M* plus one

Table 4 gives the percentages of cases falling in each of these categories, as compared to the normal group. Extratensive cases form only 28% of the total group (comparable to the 26% of the normal group), whereas the coarcted and ambiequal cases, which show no prevalence in either direction, constitute together more than half of the total sample (as against only 35% of the normals). If one further takes notice of the fact that in the EB classification the extratensive group also contains a number of cases with only one *M*, it appears that the notion of color predominance in the Rorschach records of epileptics as a group finds no support in these data. A statistical check with the normal group by means of chi square (based on the raw data instead of the percentages shown in the table) indicates a difference significant at the 1% level, mostly due to the predominance of coarcted and the relative absence of introversive cases in our sample.

It has been said (1, 4, 9) that color predominance in epileptics takes the form of a

Table 4

EB Classification of 136 Epileptics and Beck's Normals

EB	Idiopathic %	Symptomatic %	All epileptics %	Normals %
Coarcted	35	33	34	12
Ambiequal	20	21	21	23
Introversive	18	17	17	39
Extratensive	27	29	28	26

Table 5

Comparison of Idiopathic and Symptomatic Epileptics

Rorschach factor	Idiopathic ($N=60$) Median	Symptomatic ($N=76$) Median	p
Determinant:			
M	.71	.63	.50
FC	.48	.68	.01–.02
CF	.44	.61	.10–.20
C	.33	.30	.50–.70
Sum *C*	.50	.64	.20
M%	33.32	42.26	.50–.70
	Number	Number	
EB class:			
Coarcted	21	25	.20–.30
Ambiequal	12	16	
Introversive	11	13	
Extratensive	16	22	

disproportion within Sum *C*, consisting of an excess of pure *C* and *CF* responses over *FC*. One glance at Table 2 shows that this observation does not hold for our group of epileptics.

Others (7, 10, 12) have pointed out that many of the pure color responses may consist in simple color naming, the so-called *Cn* responses. They occurred in 10 cases or 7% of the whole group.

In view of the fact that others (10) have reported differences in various Rorschach scores for idiopathic and symptomatic epileptics, we also compared these two groups on the main measures used in this study. The results are brought together in Table 5. The frequency distributions of the two groups do not differ appreciably, except for *FC* on which the symptomatic group scores higher. In order to represent the more typical case, and to show once more the trend toward underproductivity on either side of the EB, we chose to publish the median values in Table 5; the preceding tables contain indications about the spread of the scores.

Discussion

The findings here reported are felt to throw considerable doubt upon the generality of inferences made from earlier studies which suggested an overemphasis on color, especially on pure *C* and *CF*, in epileptics. On the other hand they confirm once more the earlier findings of low *M* in epileptic records. Our analysis has shown, moreover, that in several cases the impression of an apparent emphasis on color is based on the artifact produced by a scarcity or even total absence of *M*. For the sample of epileptics we used, then, the outstanding feature of the EB is one of underproductivity on both sides.

Since our sample differs from those of many previous workers in intellectual level and range among other things, the conclusion suggests itself that the general mental status of epileptics (which largely determines institutionalization or extramural care) is an important factor in regulating the use of color in the Rorschach test. Guirdham (4) also called attention to this factor, but he still held to the finding of an imbalance within Sum *C* consisting in a relative neglect of *FC* in favor of *CF* and *C*. This phenomenon is not typical of our group.

The idea of a typical emotional explosiveness of epileptics, which seemed to have received some corroboration from earlier studies (mainly with institutionalized patients), would thus appear in need of experimental support. And such support should preferably not come from test scores with an insufficiently defined meaning in which the clinical impression itself functions as an interpretative factor.

An intellectually average group of epileptics, such as our sample, seems to yield a variety of EB types in which constrictive tendencies predominate. The observed variability ties in well with the growing conviction among epileptologists that personality disturbances in epileptics do not follow one typical pattern, but are multifactorial with a considerable tendency toward constriction. A narrowing of the scope of functioning is found in many groups of afflicted persons and is an earmark of general neuroticism.

Since coarctation on the Rorschach test may be related to *R*, the data were also analyzed in terms of a split between a low (*R*, 0–24) and a high (*R*, 25 or more) productivity group, each compared with the analogous subdivision of the normal group. The trend of the differences was found to lie in the same direction. For *M*, the difference between epileptics and normals remained significant in the high as well as in the low productivity group; for *C*, a significant dif-

ference was found in the high productivity group.

Conclusions

1. There is a significant difference in the EB of normal subjects and of an extramural, nondeteriorated group of epileptic veterans who show underproductivity in all determinants of the EB.

2. In contrast to the bulk of previous studies on epileptics, which point to a predominance of extratensive EB's, this extramural, nondeteriorated group of epileptics yields a variety of EB types in which constrictive tendencies predominate. Generalizations about "epileptics" from the Rorschach EB become thus a hazardous undertaking.

3. In many cases the impression of an apparent emphasis on color is based upon an artifact, produced by a scarcity or even total absence of movement responses in epileptics.

4. No significant differences in the crucial determinants of the EB (*M*, pure *C*, Sum *C*) are found between idiopathic and symptomatic epileptics (both nondeteriorated, extramural, male veterans).

5. In view of earlier publications relating aberrations in the use of the Rorschach color determinants to emotional explosiveness of epileptics, one is thus forced to select one of the following alternatives:

a. One retains the clinical notion of the emotional explosiveness of epileptics, but then the unformed color determinant of the Rorschach test does not appear to be a valid measure of this trait.

b. One accepts the diagnostic meaning of unformed color in the Rorschach test as emotional explosiveness, but then the corresponding personality trait would not appear typical of epileptics.

Summary

The Rorschach EB data of 136 extramural, intellectually grossly intact epileptic veterans were compared with those of Beck's normal controls. In all determinants of the EB the epileptics showed a significant underproductivity. The majority had a coarcted or ambiequal EB, whereas in many cases the impression of an apparent emphasis on color was based on an artifact produced by scarcity or absence of movement responses. These facts contrast with repeated statements in the literature that emotional explosiveness (as measured by an extratensive EB) is a typically epileptic character trait. It is concluded that either: (*a*) all epileptics are emotionally explosive, but Rorschach's *CF* and *C* do not measure it; or (*b*) *CF* and *C* are valid indicators of emotional explosiveness, but epileptics as a group are not characterized by this trait.

Received July 20, 1954.

References

1. Altable, J. P. Rorschach psychodiagnosis in a group of epileptic children. *Nerv. Child*, 1947, **16**, 22–23.
2. Beck, S. J., Rabin, A. I., Thiessen, W. G., Molish, H., & Thetford, W. N. The normal personality as projected in the Rorschach test. *J. Psychol.*, 1950, **30**, 241–298.
3. Collins, A. L., & Lennox, W. G. The intelligence of 300 private epileptic patients. *Proc. Ass. Res. nerv. ment. Dis.*, 1947, **26**, 586–603.
4. Guirdham, A. The Rorschach test in epileptics. *J. ment. Sci.*, 1935, **81**, 870–893.
5. Kelley, D. M., & Margulies, H. Rorschach case studies in the convulsive states. *Rorschach Res. Exch.*, 1940, **4**, 157–190.
6. Lisanski, E. S. Convulsive disorder and personality. *J. abnorm. soc. Psychol.*, 1948, **43**, 29–37.
7. Piotrowski, Z. The personality of the epileptic: the concept of the epileptic personality. In P. H. Hoch & R. P. Knight (Eds.), *Epilepsy, psychiatric aspects of convulsive disorders.* New York: Grune & Stratton, 1947. Pp. 89–108.
8. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing.* Vol. 2. Chicago: Year Book Publishers, 1946.
9. Rorschach, H. *Psychodiagnostics.* Berne: Huber, 1942.
10. Stauder, K. H. *Konstitution und Wesensaenderung der Epileptiker.* (Constitution and character changes in epileptics.) Leipzig: Georg Thieme, 1938.
11. Stotsky, B. A. A comparison of remitting and nonremitting schizophrenics on psychological tests. *J. abnorm. soc. Psychol.*, 1952, **47**, 489–496.
12. Weiszenfeld, F. P. Der Rorschachsche Formdeuteversuch als Hilfsmittel zur Differentialdiagnose zwischen genuiner Epilepsie und Uebererregbarkeitsepilepsie. (The Rorschach test as an aid in the differential diagnosis between genuine epilepsy and "overexcitement" epilepsy.) *Z. ges. Neurol. Psychiat.*, 1941, **171**, 321–336.
13. Zimmerman, F. T., Burgemeister, B., & Putnam, T. Intellectual and emotional makeup of the epileptic. *Arch. Neurol. Psychiat.*, 1951, **65**, 545–556.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

A Note on Some Projected Familial Attitudes Associated with Rorschach Movement Responses

Jerome L. Singer
Franklin D. Roosevelt VA Hospital, Montrose, New York
and Daniel A. Sugarman
Teachers College, Columbia University

This study represents a partial replication of aspects of an investigation by Shatin (5) which involved comparison of Rorschach and Thematic Apperception Test variables. Shatin reported that a significant association emerged between the number of Rorschach human movement responses (*M*) and the designation of parental figures on the TAT as "benign or nurturant." His subject sample consisted of 90 psychoneurotic patients. Recently, an increasing body of experimental data has tended to support Rorschach's empirical association of *M* responses with motor inhibition or delaying capacity, on the one hand, and fantasy or imaginativeness, on the other (2, 7, 8). Shatin's finding offers a clue to the origin of this capacity for delay presumably tapped by Rorschach's *M*. Since individuals who form numerous *M* responses tend also to describe parental figures as more benign, one can speculate that individual differences in fantasy or delaying capacity may have their origin in differing patterns of parent-child relationships. Perhaps the child who has at least some sense of security in his dependent relationship with one or both parents may be more capable of restraining motor response directed toward gratification of needs during the temporary absence of the parent. Such restraint might be accomplished by fantasied identification with the activity of the adult. In the case of the relatively consistent and benign parents the assurance of their return may reinforce the delaying capacity and the employment of fantasy as a temporary substitute. However tentative such a formulation, it seems worth while to attempt at least to confirm Shatin's finding using different subjects and another nosological group.

The principal hypothesis of the present study is that schizophrenic adults differing in production of Rorschach *M* differ predictably in their descriptions of parent-child relationships and parent identifications in Thematic Apperception Test stories. Specifically, it is hypothesized that patients designated as members of a High-*M* group attribute relatively more positive qualities to parental figures, identify adult figures more often as parents, and ascribe more positive features to the interpersonal relationship of parent and child than do Low-*M* patients.

Procedure

Both Rorschach and TAT protocols were obtained from 60 hospitalized veterans diagnosed as schizophrenics. The *S*s were divided into two groups, High-*M* and Low-*M*, using the production of two or more *M* as the cut-off point. The two groups did not differ significantly in other Rorschach scores nor in pertinent background factors, including age, IQ, diagnostic group, duration of illness, and educational experience. A further description of this sample was presented in connection with a previous study (7).

To insure relative objectivity all TAT protocols were analyzed in accordance with a scheme described in previous studies (3, 6). Only Cards 1, 2, 6BM, and 7BM of the TAT were included. The rating of the protocols was

carried out by the authors independently without knowledge of *Ss*' Rorschach group. Since scoring called only for judgments of the presence or absence of certain qualities, chi square was employed as a test of significance of the association between distributions of pertinent TAT ratings for the High- and Low-*M* groups.

Results

1. On Card 2BM ("Farm Scene"), High-*M* *Ss* showed a slight trend in more frequently designating the right-hand female figure as "Mother." ($\chi^2 = 2.2$; $p/2 = .08$)

2. On Card 6BM ("Mother-Son"), Low-*M* *Ss* more frequently failed to designate the female figure as "Mother." ($\chi^2 = 3.2$; $p/2 < .05 > .02$)

3. Both High- and Low-*M* *Ss* did not differ in designating the "mother-figure" on Card 6BM as "cold or rebuking." Since 85% of the total sample so designated the "mother-figure," little discrimination could be expected here.

4. High-*M* *Ss* described the atmosphere of the interpersonal relationship between "mother-figure" and son on Card 6BM more often as "warm" or possessing some positive elements and less often as "cold" than did the Low-*M* *Ss*. ($\chi^2 = 5.9$; $p/2 < .01$)

5. High-*M* *Ss* more often ascribed purely positive qualities to the "father-figure" on Card 7BM than did Low-*M* *Ss*. ($\chi^2 = 4.4$; $p/2 < .03 > .01$)

6. High-*M* *Ss* described the atmosphere of the father-son interpersonal relationship more often as "warm" or possessing positive elements and less often as "cold" or "neutral" than did Low-*M* *Ss*. ($\chi^2 = 4.1$; $p/2 = .02$)

7. No differences emerged in ratings of "means-end cognizance" or in the nature of problems discussed on the four cards.

Discussion

The results in general seem to confirm the hypothesis and are in accord with Shatin's findings for a neurotic sample. Since it must be remembered that these schizophrenics present considerable discord in the interpersonal relationships depicted in their stories, the finding of differences on this dimension between High- and Low-*M* *Ss* is all the more striking. As a matter of fact, these results are of interest because they tend also to offer with a different sample further support of conclusions reached in an earlier study comparing projected familial attitudes of schizophrenics and normals (6). Although social class status was not studied here (most *Ss* in this group being of lower-middle or borderline status) the marked tendency for schizophrenics to ascribe more rejection and negative characteristics to parents is borne out again in this sample. Little change in percentages from those reported for the previous sample of 60 schizophrenics emerged.

The generally negative view of parental figures appearing in TAT protocols of the schizophrenics is modified in part in the records of patients with two or more Rorschach *M* responses. Pending further verification of this result with other samples and normal groups, it seems premature to draw any final conclusions. Nevertheless, an avenue of exploration into possible genetic factors involved in the *M* response and the general dimension of motor inhibition and fantasy seems to beckon. These findings suggest the possibility that some persons have developed long-standing positive fantasies of parental figures which sustain them during periods of delayed gratification. The results reported here are consistent with views expressed by Mowrer (4) and by various psychoanalytic theorists (1) that socialization or the evolution of the reality principle in ego development is related to incorporation of parental attitudes through positive identifications with adults.

Summary and Conclusions

This investigation was designed to test with a different sample and nosological group the finding reported by Shatin (5) that neurotic *Ss* with relatively numerous Rorschach *M* tended to portray parental figures in TAT stories as "benign or nurturant." A sample of 60 hospitalized schizophrenic adult males was divided into High- and Low-*M* groups. Their protocols for TAT Cards 1, 2, 6BM, and 7BM were rated in categories relating to identification of parental figures and positive

or negative characteristics ascribed to parents or to the atmosphere of the parent-child relationship. Results revealed that High-*M* *Ss* tended more frequently to identify older female figures as mother and generally ascribed either somewhat more positive or less negative characteristics to fathers or to the interpersonal relationships between parents and children in these cards. The marked trend for these schizophrenic *Ss* to perceive mother figures as cold or rejecting and to ascribe negative characteristics generally to parent-child relationships in their stories confirms an earlier finding with another sample of patients (6). Despite this general assignment of negative traits to parental figures, the emergence of somewhat more positive parental characteristics in the stories of High-*M* *Ss* confirms Shatin's results. It is suggested that a fruitful approach to the understanding of the genesis and evolution of the Rorschach *M* and the capacity to delay overt response may involve exploration of the atmosphere of early parent-child relationships.

Received July 8, 1954.

References

1. Lair, W. S. Psychoanalytic theory of identification. Unpublished doctor's dissertation, Harvard Univer., 1949.
2. Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses: a test of the sensory-tonic theory. *J. Pers.*, 1953, **21**, 400–410.
3. Mitchell, H. E. Social class and race as factors affecting the role of the family in Thematic Apperception Test stories of males. Unpublished doctor's dissertation, Univer. of Pennsylvania, 1950.
4. Mowrer, O. H. Identification: a link between learning theory and psychotherapy. In *Learning theory and personality dynamics.* New York: Ronald, 1950. Pp. 573–616.
5. Shatin, L. Rorschach adjustment and the Thematic Apperception Test. *J. proj. Tech.*, 1953, **17**, 92–101.
6. Singer, J. L. Projected familial attitudes as a function of socioeconomic status and psychopathology. *J. consult. Psychol.*, 1954, **18**, 99–104.
7. Singer, J. L., & Herman, J. Motor and fantasy correlates of Rorschach's human movement responses. *J. consult. Psychol.*, 1954, **18**, 325–331.
8. Singer, J. L., & Spohn, H. E. Some behavioral correlates of Rorschach's experience-type. *J. consult. Psychol.*, 1954, **18**, 1–9.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Relationship Between Diagnostic Category and Deviant Verbalizations on the Rorschach

William T. Powers and Roy M. Hamlin
Western Psychiatric Institute, University of Pittsburgh

In a study of pathological thinking in the Rorschach, Watkins and Stauffacher (6) compared 25 normals, 25 neurotics, and 25 schizophrenics on the basis of their "delta scores." The authors used a modification of Rapaport's types of deviant verbalizations on the Rorschach, which were felt by Rapaport to be more or less indicative of the presence of schizophrenia, and assigned quantitative values to various types of deviant responses. Watkins and Stauffacher obtained these values by having each author arbitrarily and independently assign, to each class of responses, a value which he felt to be appropriate on the basis of his experience. The specific values ranged from .25 for mildly pathological responses, through .50 for moderately pathological responses, to 1.00 for very pathological responses. Having developed this scoring procedure, they then proceeded to score, independently, the 75 Rorschach protocols used in their study. In order to equate for different length of protocols, they divided the total score for each protocol by the total number of responses in the protocol. In each case, this figure was then multiplied by 100 to obtain the delta score for the protocol. The average delta scores for the normal, neurotic, and psychotic groups were respectively 2, 4, and 18; and *t*-test comparisons significantly differentiated the normal group from the neurotic group and the neurotic group from the psychotic group at the .01 level of confidence. Watkins and Stauffacher report a scoring reliability coefficient of .78, highly significant, which is based on total scores for all 75 records.

The present paper reports results obtained with the method of scoring developed by Watkins and Stauffacher. The outside criterion in this case consisted of five clinical groups, two of them nonschizophrenic, and the other three schizoid or schizophrenic. A specific point of interest in planning the present study was the relative efficiency of Watkins and Stauffacher's scores as compared with scores of formal Rorschach elements, "sign" scores, and clinicians' judgments. Brief comment will be made on this point in the Discussion.

The hypothesis tested in this study was: Watkins and Stauffacher's scores show a significant correlation with the five clinical groups, representing degrees of psychological disorganization.

Method

Fifty subjects were used, 10 in each of five groups. The first group consisted of 10 socially adjusted subjects. Each of the other four groups may be labeled with a diagnosis. However, diagnostic label was never accepted as a sufficient criterion for inclusion in a group. For example, group 4 consisted of 10 patients with a diagnosis of paranoid schizophrenia. Each case record was studied carefully to insure that these 10 subjects met the following criteria: agreement on diagnosis over a reasonable period of time; definitely elicitable delusions; ideas of reference or hallucinations; a history of inadequate sexual adjustment; and a record of instances in which the subject indulged in bizarre behavior. The paranoid schizophrenic group was differentiated from the catatonic schizophrenic group by the absence of mutism, posturing, and other catatonic symptoms, with these factors in turn operating as cri-

teria for inclusion in the catatonic schizophrenic group.

Each of the other four groups, including the normals, was screened with equal care and, of course, was selected for inclusion in the study before taking the Rorschach. This procedure insured somewhat more homogeneous and meaningful clinical categories than the uncritical use of diagnostic labels. A second purpose was to minimize contamination; a patient may be diagnosed as a paranoid schizophrenic partly because of his Rorschach. Every effort was made to include in the criteria for a group only acceptable clinical evidence and observation that could hardly come from the Rorschach itself.

The five groups were:

1. Socially adjusted subjects
2. Anxiety neurotics
3. "Latent" schizophrenics
4. Paranoid schizophrenics
5. Catatonic schizophrenics

Essentially, these five groups were selected to represent a rough continuum of "schizophrenic" disorganization, in line with Rapaport's description of deviant responses as characteristic of schizophrenia. However, a neurotic group was included, since Watkins and Stauffacher found that neurotic subjects obtained higher deviant response scores than normal subjects. The inclusion of a neurotic group does not imply any assumption of basic similarity between neurosis and schizophrenia. The hypothesis tested was that these deviant response scores are related to levels of psychological disorder or disorganization.

The catatonic schizophrenics were placed last in the list of five groups on the basis of consultation with psychiatrists, suggesting that they might show more disorganization in terms of deviant responses than paranoid schizophrenics.

The scoring used in this study followed Watkins and Stauffacher's procedure in all respects. In the present study, each Rorschach protocol was typed, and information as to age, sex, and education of the subject was given at the top of the protocol. In the scoring of each record, then, it seems apparent that the score assigned to any given response by a judge may be influenced by the opinion which the judge has concerning the previous responses which he has scored in this same record. In other words, each response was not evaluated *in vacuo* but in the context of all other responses in a given protocol since the scorer always had the total Rorschach before him. This may possibly be a weakness in the method of scoring, but does bring the procedure closer to a "global" approach.

Results

Validity. The mean delta scores which Watkins and Stauffacher obtained for their normal, neurotic, and psychotic groups were, respectively, 2, 4, and 18. In the present study, the mean delta scores for the normal, neurotic, latent schizophrenic, paranoid schizophrenic, and catatonic schizophrenic groups were, respectively, 6, 11, 21, 21, and 34. It is apparent that the mean delta scores in the present study are higher than those reported by Watkins and Stauffacher. The order of groups was exactly as predicted, although the difference between the latent schizophrenic and paranoid schizophrenic groups was minimal. An analysis of variance design was used and the *F* ratio was 6.2. An *F* ratio of 3.8 is significant at the .01 level for 4 and 45 degrees of freedom. An epsilon coefficient, or corrected eta, was also com-

Table 1

Number of Cases Falling in the Scoring Categories Indicated

Delta scores	Normals	Anxiety neurotics	Pre-schizophrenics	Paranoid schizophrenics	Catatonic schizophrenics
Above 50				1	1
45–49					
40–44			1	1	2
35–39			1		2
30–34					1
25–29			1		
20–24			1	2	
15–19		3	1	3	2
10–14	1	1	4	1	1
5–9	5	6	1	2	1
0–4	4				
Mean	5.62	11.40	20.53	21.00	33.50
SD	2.71	4.64	11.59	13.50	21.57

puted. This value was .65, a highly significant figure.

Table 1 shows the distribution of the subjects in the five groups with respect to their delta scores. The mean and standard deviation of each group are also indicated. Inspection of this table suggests cutoff points at scores 10 and 20. Only one subject, or 10%, of the normal group scored higher than 10, whereas 80% to 90% of each of the psychotic groups scored above 10. However, 40% of the neurotic group obtained scores above this cutoff point. A cutoff point at score 20 shows no normal or neurotic subjects obtaining higher scores, whereas 40% to 60% of each psychotic group scored above this point. Table 2 indicates the percentage of cases in each group falling above the selected cutoff scores.

Table 2

Percentage of Cases in Each Group Falling Above the Indicated Delta Scores

Delta score	Normals	Anxiety neurotics	Preschizophrenics	Paranoid schizophrenics	Catatonic schizophrenics
Above 20	0	0	40	40	60
Above 10	10	40	90	80	90

Any practical value that such cutoff points might have is, however, contingent upon strict consistency of scoring. The delta scores in the present study show a constant trend toward relatively higher values than those reported by Watkins and Stauffacher. This difference in scores is reflected in their suggested cutoff points at scores 5 and 10. Careful standard training of scorers is, of course, indicated, but there is also apparent need for further refinement and specification of scoring categories.

Scoring reliability. In Watkins and Stauffacher's study, the reliability of scoring coefficient, based on total scores, for all 75 records was .78, a highly significant figure. In the present study, only one author scored all 50 Rorschachs. Fifteen protocols were selected at random and scored by the second author to provide an estimate of reliability. Each scorer worked independently, after careful review of Watkins and Stauffacher's definitions and examples of the various scoring categories. A product-moment correlation coefficient was computed on the basis of the total scores assigned to each of these 15 Rorschachs. The obtained value of .88 supports the result obtained by Watkins and Stauffacher.

When the specific responses in each protocol are considered, however, the results are less satisfactory. The total number of responses in all 15 protocols amounted to 441. Of this total, one author scored 133 responses while the second scored 142 responses. The total number of jointly scored responses was 87, with the percentage of agreement slightly over 60. In the individual protocols the percentage of agreement ranged from 50 to 86.

The results are even less satisfactory when the categories to which the jointly scored responses were assigned are considered. Seldom was a jointly scored response assigned to exactly the same category by both judges. These results suggest that the scoring categories should be more adequately delimited, and those categories which are extremely unreliable or which serve little function in differentiating degrees of pathology should be eliminated or at least modified.

Watkins and Stauffacher have considered the problem of consistent scoring in some detail, and their data point to the same difficulties suggested here. One important indication from their results should be repeated: the reliability of scoring increases consistently with degree of psychological disorganization. Specifically, they report separate scoring reliabilities for their 25 normal, 25 neurotic, and 25 psychotic subjects. These coefficients are, respectively: .04, .47, and .91. Although the increasing reliability is related simply to the increasing magnitude of scores in the more pathological groups, two other factors should be considered. First, the "mildly deviant" responses, as specified at present, may be well nigh impossible to score with any suitable degree of reliability. Second, in both studies, the scorers worked with the entire protocol before them, not with one strictly isolated response at a time. The higher reliability for the more pathological groups may reflect global estimates of the Rorschachs, guided

by specific attention to one response at a time. The tendency, or constant error, of the scorers in the present study to obtain higher values than those reported by Watkins and Stauffacher may be related to these considerations.

In spite of these scoring problems, it is nevertheless clear that encouraging coefficients of correlation were obtained for both reliability of scoring and for validity when total scores were used in conjunction with a correction factor for variability in number of responses, and when each independent worker scored all records considered in any given treatment.

Discussion

The specific focus of interest in undertaking the present study was the relation of Watkins and Stauffacher's method to other treatments of Rorschach data. The delta scores used here are superficially similar to Rorschach sign or determinant scores, or to judgments of small units of Rorschach material. Actually the delta scores represent a method midway between the sign approach and the use of global judgments. Hamlin (4) has previously reported on a series of 10 studies using variously the sign, small unit judgment, and global judgment approaches. These studies suggested that small, mechanically scored or judged units tend to give negative results when correlated against an outside criterion, and also that global judgments tend to give negative results when the material to be judged is too complex. The best results seem to be obtained in those situations where larger units than signs are used, and where samples of behavior sufficient to represent some degree of organization are used so that a clinical judgment is possible rather than a mechanical scoring without a rationale.

The method followed with the delta scores in the present study incorporates several features of interest in this respect. First, the scorer always worked with the total Rorschach before him. Second, his attention was directed to a small unit of projective behavior, a single Rorschach response, at a time. Each score or judgment was made on a small unit of material, but in the more global context of a total Rorschach. Third, the scoring involved an underlying rationale based on Rapaport's formulations. Thus, the scorer had a series of feasible tasks, each in a larger configurational context, and knew what he was doing in terms of a meaningful rationale.

The results suggest that these conditions contribute to an encouraging degree of correlation with an outside criterion of psychological disorganization. For example, as a rough indication, the obtained epsilon coefficient of .65 compares favorably with the values of .47 and .50 cited by Hamlin (4) from two studies (1, 3) where the outside criterion groups (representing degrees of psychopathology) and the variable judged (overall adjustment) were somewhat similar. The correlation of .68 cited from another study (2), where Wechsler-Bellevue IQ was the outside criterion, is not significantly higher.

The general questions underlying this and the other studies mentioned have been: under what conditions can clinicians make valid judgments from projective material? and how do successful judges proceed in making predictions? The possible practical value of such scales as Watkins and Stauffacher's for routine clinical use is questionable. Techniques like the Rorschach will continue to be used as clinical tools without being reduced to an intricate variety of objectively scored scales. The study of judgments and of conditions necessary for successful judgments may, however, clarify the efficiency of projective procedures.

Summary

The hypothesis tested in this study, that Watkins and Stauffacher's delta scores will show a significant correlation with the five diagnostic groups, was upheld. The *F* ratio of 6.2 and the epsilon value of .65 were both significant at the .01 level, as was also the .88 scoring reliability coefficient obtained from a comparison of total scores assigned by the two authors to 15 jointly scored protocols.

The significant results were in keeping with the results obtained by Cummings (3), Albee and Hamlin (1), and Watkins and Stauffacher (6) in similar studies, and the meth-

ods used in the present study resembled the methods used in those studies.

In all of these studies, as opposed to less successful judgment studies, the scorer or judge dealt with small units of material and made judgments based on a specific rationale. The procedure of allowing the judges to have the complete Rorschach protocols before them while evaluating the specific responses gives the added advantage of allowing the judge to gain an over-all impression of the record, while at the same time his forced consideration of each individual response insures that his final judgment of the protocol will by no means issue solely from his opinion of the total gestalt.

Received August 5, 1954.

References

1. Albee, G. W., & Hamlin, R. M. Judgment of adjustment from drawings: the applicability of rating scale methods. *J. clin. Psychol.*, 1950, **6**, 363–365.
2. Bialick, I., & Hamlin, R. M. The clinician as judge: details of procedure in judging projective material. *J. consult. Psychol.*, 1954, **18**, 239–242.
3. Cummings, S. T. The clinician as judge: judgments of adjustment inferred from Rorschach single-card performance. *J. consult. Psychol.*, 1954, **18**, 243–247.
4. Hamlin, R. M. The clinician as judge: implications of a series of studies. *J. consult. Psychol.*, 1954, **18**, 233–238.
5. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing*. Vol. 2. Chicago: Year Book Publishers, 1946.
6. Watkins, J. G., & Stauffacher, J. C. An index of pathological thinking in the Rorschach. *J. proj. Tech.*, 1952, **16**, 276–286.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Measurement of Anxiety in Clinical Situations[1]

Arnold H. Buss, Morton Wiener, Ann Durkee, and Marc Baer

Carter Memorial Hospital, Indianapolis

Anxiety has been measured in a variety of ways, but there is little general agreement among measures. A possible source of this disagreement is the methodological problems involved in evaluating anxiety. The problems encountered in measuring clinical anxiety are primarily those of definition and reliability of measurement. Because the term "anxiety" has a variety of behavioral referents, the operational definition should be explicit, and the behaviors labelled "anxious" should be specified. When anxious behavior is rated in clinical situations, there should be several judges and a statement of interjudge agreement.

Two previous studies illustrate attempts at solving these problems. Gleser and Ulett attempted to measure "anxiety-proneness," which was defined as "the present mental status, past history of anxiety, background and personality development that would indicate the occurrence of anxiety symptoms in a stressful situation" (4, p. 280). A psychiatrist and a psychologist used an eight-point scale to rate anxiety-proneness. The interrater reliabilities were .51 for patients and .30 for normals. Elizur defined anxiety as "an inner state of insecurity which may take one or more of the following forms: fears, phobias, lack of self-confidence, extreme shyness, ideas of reference, and marked sensitivity" (3, p. 248). From protocols of interviews with college students three judges made anxiety ratings on a nine-point scale. The average interrater correlation was .70.

In evaluating these two studies it may be seen that Elizur's definition of anxiety is more specific than that of Gleser and Ulett. This was evidently reflected in a considerably higher interjudge reliability in rating anxiety in normals: .70 for Elizur and .30 for Gleser and Ulett. It seems likely that further specificity in defining anxious behaviors would yield still higher interjudge reliability and would prove more useful in future research in this area. If reliability can be established, the obtained measure could be used as a criterion for validating other measures, such as anxiety questionnaires.

Purpose

The purposes of this paper were: (*a*) To attempt a clear delineation of various manifestations of anxiety and to determine if this results in higher interrater reliability; (*b*) to determine the relationship of each manifestation of anxiety to a global evaluation of anxiety; (*c*) to use the ratings of anxiety as a criterion against which to validate the Taylor Scale of Manifest Anxiety (*A* scale) (7).

Method

Subjects. The *S*s were patients in a neuropsychiatric hospital. Patients were excluded from the study for the following reasons: (*a*) out of contact, (*b*) on electroshock or insulin therapy, (*c*) low intelligence, (*d*) diagnosed organic brain damage. Seventy-three patients (33 men, 40 women) were interviewed, and 64 of these completed the *A* scale. Of the nine *S*s who failed to complete the *A* scale, several stopped in the middle and refused to go on, and the remainder complained of being ill.

Procedure. Each *S* was interviewed in a room with four psychologists, who served as judges. A pilot study had indicated that the

[1] This study was carried out with the permission and encouragement of Juul Nielson, M.D. The writers are indebted to Drs. Charles Windle and Leonard Goodstein for their critical reading of the manuscript.

presence of additional persons in the interviewing room did not markedly affect *S*'s anxiety level. Patients in this hospital are frequently seen by several members of the staff simultaneously. Each *S* was interviewed by the same psychologist, although the other three psychologists also asked questions at the end of the interview. The interview lasted approximately 15 minutes. Then the *S* was conducted to another room, and the *A* scale was administered by another staff member. Immediately after the interview, each judge rated *S* for anxiety. Each rating was made independently and without consultation with the other judges.

The interview. The interview was relatively uniform. The interviewer introduced himself and then the other three judges to *S*. At first all questions were open-ended and unstructured, e.g., "What has been bothering you?" "What happened?" "How do you feel inside?" As the interview progressed, the questions became more specific and direct, e.g., "Do you have aches and pains?" "Do you have trouble sleeping?" "Does your heart beat fast?" At the end of the interview, the interviewer asked the other judges if they had any questions to ask. Finally, *S* was given the task of subtracting sevens from 100, and the interview was ended.

The ratings. Although definitions of anxiety in psychiatric and abnormal psychology texts (1, 2, 5, 6, 8) vary, there tends to be general agreement on the manifestations of anxiety. In this study an attempt was made to use only those manifestations of anxiety on which there is general agreement. It was decided to group symptoms into categories because the probability of occurrence of any given symptom, e.g., excessive sweating, in any given *S* was small. Each category was rated on a five-point scale: a rating of 1 represented absence of or minimal manifestation, and a rating of 5 represented maximal manifestation.

The categories were divided into behavior observed and behavior reported by the patient in the situation, as follows:

A. *Observed*

1. *Distractibility.* It was found in the pilot study that distractibility per se was difficult to rate. Therefore a task was introduced, the task of subtracting sevens from one hundred. Serial subtraction is typically used in psychiatric settings as a measure of distractibility. Inattentiveness during the interview was also taken into account in rating this category.

2. *Restlessness.* This category involved the skeletal components of anxiety: agitation, tremors, and tics. It was arbitrarily decided to rate at least 4 (on the five-point scale) for the presence of tremor and to rate 5 for marked tremor.

3. *Physiological concomitants.* This category involved the autonomic components of anxiety. Four manifestations were selected: sweating, flushing, breathing disturbance, and excessive swallowing. The rating was based on both intensity and number of symptoms.

B. *Reported*

4. *Subjective feeling of tenseness.* This category referred to the patient's report of how much he was bothered or upset. The rating was based on the intensity, frequency, and chronicity of the complaint. It was arbitrarily decided that the verbalization of panic would be rated 5.

5. *Worry.* This category referred to apprehension, regardless of whether there was a basis in fact. It was arbitrarily decided to rate 2 if *S*'s worry concerned only money or leaving the hospital. The rating was based on intensity and chronicity.

6. *Somatic complaints.* This category referred to aches, pains, and peculiar feelings exclusive of actual illness, e.g., gallstones, arthritis, or heart disease, although somatic delusions were included. The ratings were based on the number, frequency, intensity, and chronicity of the complaints.

7. *Physiological concomitants.* This category involved the same manifestations as observed physiological concomitants, with the addition of palpitation. It might be expected that the observable concomitants would be the same as the reported concomitants. However, it was demonstrated in the pilot study that these two categories were not identical; several patients reported physiological concomitants without manifesting them in the interview situation.

8. *Muscular tension.* This category involved the same manifestations as observed restlessness, with the addition of a feeling of internal tension. At first glance, muscular tension may appear to be identical with subjective feelings of tenseness. However, as defined in this study, it referred to an awareness of muscular contraction that was reported as "tied up in knots," "ready to explode," "feel shaky." This may be contrasted with the reports on subjective feelings of tenseness: "nervous," "upset," "stirred up inside," "panicky."

9. *Over-all rating of anxiety.* This category refers to a global estimate of anxiety. The ratings were based on what may be termed a clinical impression of anxiety. The judges attempted to divorce this rating from any particular category or configuration of categories.

Results

The anxiety ratings of each judge were compared with those of every other judge. Thus, for each of the nine categories there were six correlations.[2] The mean correlations between judges are shown in the first two columns of Table 1. The mean correlation of .83 between judges for over-all anxiety indicates fairly high interrater reliability.

The ratings for distractibility appear somewhat more reliable than the ratings in the other categories. The rating for distractibility was based largely on the patient's performance on the task of subtracting sevens. The number of errors and the amount of time taken to perform this task are fairly objective measures, which probably accounts for the high reliability.

The ratings for observed physiological concomitants seem somewhat less reliable than the ratings of the other categories. This may be ascribed to a fundamental difficulty in observing autonomic changes. Changes in the depth and rate of breathing are discernible, but changes in sweating, flushing, and swallowing are often so minimal as to escape notice. The ratings for the remaining categories yielded mean intercorrelations that ranged from .66 to .85, indicating a moderate degree of reliability in assessing these aspects of anxiety.

The eight individual categories of manifest anxiety were selected on the basis of their expected relationships to global anxiety. In order to determine these relationships the ratings in each category were pooled [3] and then compared with the pooled ratings of over-all anxiety. The correlations, shown in the third column of Table 1, indicate that there is a definite relationship between the over-all rating of anxiety and the ratings in each individual category. While the correlations range from .47 to .72, there is little difference between adjacent correlations, and the differences may be ascribed to chance fluctuation.

[2] All correlations in this study are Pearson product-moment *r*'s; the correlations were averaged by the *z* transformation method.

[3] It seems reasonable to assume that the best estimate of each *S*'s rating for any given category is the consensus of the four judges. Therefore the ratings of the judges were averaged, and these pooled ratings were used in all subsequent comparisons.

Table 1

Intercorrelations Among Judges, and the Correlations of the Ratings with Over-all Anxiety and with the Taylor *A* Scale

Category	Intercorrelation among judges		Correlation with over-all rating	Correlation with *A* scale
	Mean	Range		
Observed				
1. Distractibility	.93	.92–.94	.47	.16
2. Restlessness	.74	.68–.80	.66	.37
3. Physiological concomitants	.56	.41–.65	.56	.22
Reported				
4. Subjective feelings of tenseness	.66	.59–.77	.65	.52
5. Worry	.76	.71–.79	.50	.50
6. Somatic complaints	.85	.81–.88	.52	.40
7. Physiological concomitants	.81	.76–.86	.72	.68
8. Muscular tension	.73	.67–.77	.68	.54
9. Over-all rating of anxiety	.83	.66–.92		.60

There were two categories of physiological concomitants, one observed and one reported. It might be expected that there would be a "halo-effect," resulting in a high degree of relationship between these categories. However, the correlation is .30, indicating a rather low degree of relationship. Evidently, *S*'s report of autonomic disturbance differed from what the judges observed in the interview situation.

Similarly, restlessness and muscular tension referred to essentially the same manifestations of anxiety, the former observed and the latter reported. The correlation between ratings in these categories is .61, indicating a definite relationship. Evidently there is a greater correspondence between observed and reported skeletal symptoms than there is between observed and reported autonomic symptoms.

Next, the pooled ratings for each category and for the over-all rating of anxiety were compared with the Taylor *A*-scale scores. The correlations are shown in the last column of Table 1. There is a low relationship between the *A* scale and the observed categories and a moderate relationship between the *A* scale and the reported categories. The corre-

lation for reported physiological concomitants was somewhat higher than that for the other categories, which suggests that there are many items of this sort in the *A* scale. The over-all rating of anxiety correlated .60 with the *A* scale, which suggests that the *A* scale measures many of the same kinds of behavior as the ratings do.

The over-all rating of anxiety was a "clinical" integration of the eight categories of anxious behavior. Integration may also be obtained by arithmetic summation of the ratings in each category; the pooled ratings for categories 1 to 8 were added to yield a single score for each *S*. The correlation between these arithmetic sums and the over-all ratings is .93, which suggests that there is little difference between arithmetic and clinical integration. The arithmetic sums of each judge were correlated with those of every other judge. The mean intercorrelation is .88, and the range is .81 to .91. The correlation between the pooled arithmetic sums and *A*-scale scores is .64. This is not significantly different from the correlation of .60 between the over-all ratings and *A*-scale scores.

Discussion

In this study clear delineation of the manifestations of anxiety resulted in reliable ratings. The average intercorrelation between judges of .83 may be compared to the correlation of .51 obtained by Gleser and Ulett (4) on their ratings of patients. With normal subjects Elizur (3) obtained an average interjudge correlation of .70, while that of Gleser and Ulett was .30. Both Elizur's definition and the present one were more specific than that of Gleser and Ulett. Thus, it may be concluded that specificity of definition increases interjudge reliability in the measurement of anxiety.

It has been assumed heretofore that the various aspects of anxiety must be integrated more or less intuitively in order to obtain a global measure of anxiety. This procedure was followed in the present study, in which an over-all clinical rating of anxiety was made by each psychologist. A likely basis for assuming the necessity of such intuitive integration is that there must be an implicit "weighting" of the various aspects of anxiety. The psychologist is expected to use his clinical judgment in deciding which aspects are more important in any given *S*. The question arises as to the advantages of clinical integration over mere arithmetic summation. The results showed an extremely high relationship between these two types of integration and that arithmetic summation yields slightly higher interjudge reliability. Apparently, clinical integration offers no advantages over arithmetic summation. Thus, it may be concluded that specific denotation of the various manifestations of anxiety, followed by summation of the ratings, yields an adequate, reliable measure of anxiety. It would appear that there is no necessity for vagueness in defining or measuring anxiety.

The manifestations of anxiety listed in most texts include both observed and reported behavior. The definition of anxiety in this study included both aspects. This is reflected in the relationship between each type of behavior and the over-all clinical rating. The correlations between the ratings for the observed categories and the over-all clinical rating were of approximately the same order as the correlations between the ratings for the reported categories and the over-all rating. Thus the over-all rating did take into account both types of behavior.

Since the clinical evaluation of anxiety conformed to a consensual definition of anxiety and proved to be reliable, it was used as a criterion measure. The *A* scale showed a moderately high relationship to both the arithmetic summation and the over-all clinical estimate of anxiety. The *A* scale showed a much higher relationship to the reported categories than to the observed categories. This was expected because the *A* scale consists only of self-ratings, i.e., reported manifestations of anxiety. Anxious behaviors observed by others are omitted. To the extent that the *A* scale omits items that relate to observed measures of anxiety, it does not completely measure anxiety, as defined by psychiatric and abnormal psychology texts. However, the *A* scale's relationship to the criterion suggests that it is a fairly adequate measure of anxiety.

It should be possible to construct an anxiety inventory with an extremely high rela-

tionship to an external criterion. This might be done by performing an item analysis on the *A* scale, using the clinical ratings to select high-anxious and low-anxious subjects. Such an inventory would provide a quantitative estimate of anxiety, as defined clinically. The advantage over present-day anxiety inventories, which have only face validity, is obvious.

Summary

This study attempted to evaluate anxiety (as defined by textbooks) in a clinical situation and relate this measure to anxiety as measured by the Taylor Scale of Manifest Anxiety. Seventy-three neuropsychiatric patients were interviewed, and 64 of these completed the *A* scale. Four judges rated the *Ss* for various aspects of anxiety on the basis of the interview material. The results were as follows:

1. Specific denotation of the manifestations of anxiety resulted in reliable ratings.

2. "Clinical" integration of anxious behavior to yield a global measure of anxiety correlated .93 with mere arithmetic summation of the various manifestations.

3. The correlation between the over-all rating of anxiety and *A*-scale scores was .60.

Received July 29, 1954.

References

1. Cameron, N. *The psychology of behavior disorders.* Boston: Houghton Mifflin, 1947.
2. Cattell, R. B. *Personality.* New York: McGraw-Hill, 1950.
3. Elizur, A. A content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Res. Exch.*, 1949, 13, 247–283.
4. Gleser, Goldine, & Ulett, G. The Saslow Screening Test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, 8, 279–283.
5. Henderson, D., & Gillespie, R. D. *A text-book of psychiatry.* London: Oxford Univer. Press, 1950.
6. Noyes, A. P. *Modern clinical psychiatry.* Philadelphia: Saunders, 1953.
7. Taylor, Janet A. A personality test for manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 41, 285–290.
8. Weiss, E., & English, O. S. *Psychosomatic medicine.* Philadelphia: Saunders, 1949.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Cross Validation of the Organic Brain Pathology Scale for a Test of Symbol Arrangement[1]

Theodore C. Kahn

USAF School of Aviation Medicine, Randolph Air Force Base, Texas

The validation of the organic brain pathology scale of a technique which requires the subject to arrange 15 plastic objects several times on a segmented strip appeared in 1951 (1). The objects used in the test vary in color, size, and thickness, and their shapes represent easily recognized symbols which are common in our culture. The administration takes approximately 15 minutes and the scoring a like amount of time. In the original study 50 patients with a diagnosis of organic brain pathology were matched with a non–brain-damaged group. In that study 197 variables were statistically evaluated and 99 of these were found to be significant. The difference of the means yielded a t ratio of 17.6 when the distributions of the scores of the two groups were compared.

Thirty patients in a state mental hospital who had a diagnosis of organic brain pathology and whose medical histories roughly corresponded to the previous patient group were matched in age, IQ, educational level, race, sex, religion, and national origin with a non–brain-damaged group of like size. The screening of the control group was carried on under the supervision of a psychiatrist. Factors which were not equated and which were significantly different in the two groups were the number of subjects who had been veterans, the number who lived in rural versus urban environment, occupational level, and marital status.

[1] An extended report of this study may be obtained without charge from Major T. C. Kahn, USAF School of Aviation Medicine, Randolph Field, Texas, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4482 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

All of the scores of the new organic group fell within the range of scores previously established by the first organic group with the exception that one patient scored lower (away from normal) than had any of the organics in the previous testing. Likewise, the scores of the new non–brain-damaged group fell within the range of scores of the first non–brain-damaged group except that the score of one subject, who was over 75 years old, fell somewhat outside of this range into the direction of the organic group. The mean score of the new organic group was -226.2 with a σ of 110.9; that of the non–brain-damaged control group was $+34.6$ with a σ of 65.2. The difference of the means of the scores of the two groups yielded a t ratio of 11.1. Since the analysis of variance produced a significant F ratio, the medians were compared by chi square and the difference was found to be significant at better than the .001 level.

Cutoff scores were established for the four groups in the two studies. If three subjects over 70 years old are excluded from the two non–brain-damaged groups, a cutoff score of -54 would separate 92 per cent of the first organic group, and 100 per cent of the second organic group, from all the non–brain-damaged subjects in both groups. Even when these three low scoring subjects are included, a cutoff score of -138 separates all the non–brain-damaged subjects in the two studies from 72 per cent of the first organic group and from 83 per cent of the second organic group.

Brief Report
Received January 12, 1955.

Reference

1. Kahn, T. C. An original test of symbol arrangement validated on organic psychotics. *J. consult. Psychol.*, 1951, 15, 439–444.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Digit Symbol Performance as a Function of Increasing Levels of Anxiety

Joseph D. Matarazzo and Jeanne S. Phillips

Division of Psychosomatic Medicine and Department of Neuropsychiatry, Washington University School of Medicine

In two earlier studies (2, 3) we reported a significant curvilinear relationship between performance in a timed learning situation and anxiety level as measured by the Taylor Scale of Manifest Anxiety (*A* scale) (6, 7), and a significant rectilinear relationship between anxiety level and intelligence as measured by a timed intelligence test. The present study was designed to extend these findings by employing a timed learning task, the Digit Symbol substitution test, which in the past has been employed both as a measure of learning and as an intelligence subtest. With the traditional procedure the 67-item Digit Symbol subtest of the Wechsler-Bellevue (10) was extended to 175 items. These were printed on one side of a sheet of paper. The time limit was increased from 1½ minutes to 3 minutes. The directions used were the standard ones.

In a group administration, a total of 119 *Ss* was used in the experiment. This number included an entire class of first-year medical students ($N = 81$) and approximately one-half of the sophomore class ($N = 38$). It was learned before the experiment that one student had had previous experience with the digit symbol task and he was excluded from the study. The *A* scale was administered to all *Ss* immediately following the learning task.

The hypothesis to be investigated was that digit symbol performance would have a curvilinear relationship to increasing levels of anxiety. On the basis of the results of the earlier stylus maze study (3), we predicted a curvilinear relationship based on the following criteria: (*a*) a U-shaped curve as determined by visual inspection, (*b*) a higher value of epsilon than *r* and a correspondingly significant value for the chi-square test of linearity of regression, and (*c*) a significant difference between the means of groups 1 and 2, 1 and 3, 1 and 4, 1 and 5, 3 and 6, and 3 and 7. These differences in means for specific groups (criterion *c*) were predicted on the basis of the significant differences obtained for the time criterion in the original study. It was realized at the outset that these predictions were contingent upon the way in which *Ss* of the present study distributed themselves in groups according to their Taylor raw scores. If an insufficient number of *Ss* fell in any group, predictions about that group would necessarily no longer be valid.

Results

On the basis of their Taylor raw scores the 119 *Ss* were assigned, as in our previous studies, to groups on a scale of increasing anxiety level. Individuals with raw scores of 1–5 were given an anxiety rating of 1, while those with raw scores of 6–10, 11–15, 16–20, etc., were given ratings of 2, 3, 4, etc., respectively. While the over-all distribution of Taylor raw scores for *Ss* of this study (ranging from 1–35, mean of 12) was similar to that reported by Taylor (7) and also to that found with the maze and intelligence data, the number of *Ss* in each subgroup (especially the groups at the high end) differed slightly from our previous findings. Thus, only 11 *Ss* earned Taylor raw scores above 21 (five from 21–25, four from 26–30, and two from 31–35). In an effort to avoid the problems imposed by these small *N*'s, the 11 *Ss* were grouped together. This left five groups: group 1 included raw scores 1–5, and individuals

with raw scores of 6–10, 11–15, 16–20, and 21–35 were placed in anxiety groups 2, 3, 4, and 5. The number of *Ss* in each of these five groups from low to high was 18, 41, 34, 15, and 11, respectively.

Before the mean digit symbol performance of each of these five groups was computed, criterion *c* of the original prediction was changed to conform to these findings. The prediction was now changed to read: based on the results of the maze study, the following groups will be expected to differ significantly: 1–2, 1–3, 1–4, and 1–5. Although groups 3–5 did not differ significantly in the earlier study (groups 3–6 and groups 3–7 did), it is possible that 3–5 will differ in this study since group 5 now includes *Ss* with raw scores ranging from 21–35, and thus overlaps groups 5, 6, and 7 of the previous study.

The mean digit symbol scores for each of the five groups are shown in Fig. 1. The values from low to high anxiety are: 101.8, 109.7, 109.8, 108.8, and 106.4. These values and visual inspection of the curve confirm criterion *a* of our hypothesis, i.e., that the function is a U-shaped curve. Two aspects of the curve are worth noting: (*a*) as with the time measure of the maze data, the largest mean difference is found between groups 1 and 2, and (*b*) the failure of the curve to make a complete U at the high end. This latter may be a function of the unfortunate restriction in the number of groups at the high end.

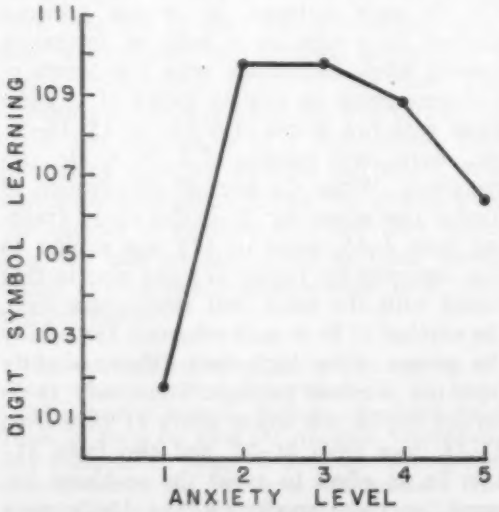


Fig. 1. Mean digit symbol score for each of the five groups ranging from low anxiety (group 1) to high anxiety (group 5).

Since the data in Fig. 1 represent only the means of the learning scores in each of the five groups, the confirmation of the hypothesis, by visual inspection alone, is not possible. This is especially true since there was considerable variability in learning score from individual to individual within each of the five groups (the sigma was roughly of the order of 15 for each of the five groups).

Criteria *b* and *c* were next applied. As in the two previous studies (2, 3) the *F*-test value of 1.02 was not significant. The reasons for this are discussed in the earlier studies and involve the fact that there may be a *functional* relationship between the two variables (digit symbol performance and anxiety level) and that the five anxiety groups are not "independent" but ordered in a sequential relationship to each other. This relationship may increase the "within" variance while not influencing the "between" variance. This whole problem, as well as the permissibility of using *t* tests with data of this type, in the absence of a significant *F*, has been discussed by Webb and Lemmon (8, 9).

Analyses of the differences between the means for the various groups revealed that two of the four differences predicted were statistically demonstrated: groups 1–2 and 1–3 were each found to differ significantly at the .03 level of confidence. The remaining two differences, between groups 1–4 and 1–5, were not demonstrated. The difference between groups 1–4 was just short of the .05 level, however.

The additional prediction, that groups 3–5 might possibly differ significantly, did not hold. Despite the relatively moderate mean difference between these two groups, shown in Fig. 1, the intragroup variability was such that the difference was not statistically significant. This finding is unfortunate for the hypothesis since it was the only difference (due to the restricted number of *Ss* at the high end) predicted between the high and middle ranges of the anxiety continuum. Consequently, criterion *c* for the establishment of a curvilinear relationship between digit symbol performance and increasing levels of anx-

iety was not confirmed. That it would have been confirmed if we had had two additional groups at the high end is a matter simply for conjecture. The demonstrated superiority of *Ss* in the middle range of anxiety, groups 2 and 3, relative to *Ss* in the low anxiety group, group 1, is nevertheless a significant finding.[1]

As was pointed out in the two previous studies, the most meaningful determination of the relationship between the two variables under discussion is the use of correlational techniques (criterion *b*). These methods allow determination of the *strength* of the relationship shown in Fig. 1. Both the correlation ratio, eta (1, pp. 259–261), and the unbiased correlation ratio, epsilon (4, pp. 319–324), can be computed from the analysis of variance data.

The values of both of these measures (.20 and .00, respectively) did not reach statistical significance and therefore do not differ from zero. This finding, which is unlike the data from the maze study, is surprising in view of the similarity between our present Fig. 1 and the earlier curve relating time to learn the maze with anxiety level. The value of the Pearson *r* was found to be an insignificant .05. Since neither correlation ratio is statistically significant, it was not possible to apply the chi-square test of linearity of regression (4, p. 319) in order to determine whether or not the true function deviates from a straight line.

Discussion

One is thus forced to conclude, within the limits of these data, that the hypothesis of a curvilinear relationship between digit symbol learning and increasing levels of anxiety—with optimal learning in the middle range—is not borne out. The hypothesis is supported to the extent that the relationship between the low anxious and middle anxious groups was demonstrated, but the equally important relationship (for the hypothesis) between the middle group and the high anxious group, despite the trend in means, was not borne out. The failure to demonstrate with the present task even the low-order, but nevertheless significant, correlations attained in the earlier studies, in view of the mean differences shown in Fig. 1, poses a question regarding the "meaning" of these same differences and also of the numerous similar ones reported in other studies using the *A* scale.

That there are, in fact, true differences in conditioning and learning abilities of *Ss* who differ in Taylor anxiety score has been demonstrated many times.[2] The present study adds to this growing literature. However, the equally important finding of low correlations (when these have been computed) between anxiety level and performance in various situations implies what has long been known, that factors other than anxiety, or intelligence for that matter, are involved in learning ability. Accordingly, correlational techniques would appear necessary in the design of future studies utilizing the Taylor scale.

Summary

On the basis of previous work, a curvilinear relationship was predicted between Digit Symbol performance and score on the Taylor Scale of Manifest Anxiety. A total of 119 *Ss* was divided into five groups ranging from low to high anxiety. The prediction was borne out in part, the performance of the middle groups being superior to that of *Ss* in the low anxious group but, despite a trend in the predicted direction, the hypothesis was not confirmed for the middle and high anxious groups. The *strength* of the relationship, however, was found by correlational techniques to be an insignificant one. A methodological question regarding the use of the Taylor scale was raised, i.e., the need to include correlational methods of analysis in future studies because of the low-order relationships obtained in studies to date.

Received July 21, 1954.

References

1. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.

[1] Consistent with the finding of our previous study, the *Ss* comprising the five groups of this study showed no differences in mean intelligence score as determined by the CVS abbreviated intelligence scale.

[2] It is of interest in view of the hypothesis of our present study that Spence and Taylor have recently reported (5, p. 270) finding a *curvilinear* relationship between conditioning and anxiety level in two unpublished studies.

2. Matarazzo, J. D., Ulett, G. A., Guze, S. B., & Saslow, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.*, 1954, **18**, 201–205.
3. Matarazzo, J. D., Ulett, G. A., & Saslow, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, in press.
4. Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1940.
5. Spence, K. W., & Taylor, Janet A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.
6. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
7. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
8. Webb, W. B., & Lemmon, V. W. A qualification in the use of analysis of variance. *Psychol. Bull.*, 1950, **47**, 130–136.
9. Webb, W. B., & Lemmon, V. W. A sequel to the notes of Patterson and Diamond. *Psychol. Bull.*, 1952, **49**, 155.
10. Wechsler, D. *The measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Discriminative Power of the Strauss-Lehtinen Battery in Adult Mental Patients[1]

A. A. Rafi
Netherne Hospital, Surrey, England

So far no experiments have been reported with adult subjects that duplicate the researches of Strauss and Lehtinen (1) with children. These authors emphasize the capacity of their battery to discriminate normals, endogenous mental defectives, and exogenous mental defectives but offer no quantitative evidence.

The present investigation aimed at determining whether the battery discriminated to a statistically significant extent adults of analogous categories, i.e., bright and dull mental patients without brain damage, mental defectives, and brain-damaged persons. Two statistical techniques were employed: analysis of variance by double classification to establish the general effect of diagnostic and sex differences on the separate tests of the battery, and discriminant function analysis to ascertain the efficiency with which the whole battery distinguished the particular categories.

Method

The test material, instructions, and testing procedures were modeled on those described by Strauss and Lehtinen (1). The Marble-Board, Figure-Background, and Tactual-Motor tests were quantified in the following manner:

Marble-Board Test (six patterns—maximum score 18)

Score

3 constructive performance of pattern;
2 orderly continuous sequence along outline of pattern;
1 discontinuous, sporadic performance with outline similar to pattern;
0 discontinuous, incoherent, sporadic performance with outline dissimilar to pattern.

Figure-Background Test (nine items—maximum score 27)

Score

3 correct naming of object with no reference to background;
2 noting presence of object in vague, imprecise manner as "round thing," "long thing";
1 mentioning both object and background but imprecisely, e.g., "scribbling with something round in the middle";
0 describing background, not mentioning object.

Tactual-Motor Test (three items—maximum score nine)

Score

3 figure only;
2 figure with contour of board;
1 figure and background;
0 background only.

The behaviors scored 0 are stated to be characteristic of brain-injured children (1).

The Dot-Perseveration Test was also administered, but perseveration appeared too rarely for adequate quantification in measurement terms and when treated enumeratively the differences were not statistically significant. This test was dropped from the battery.

All the subjects were inpatients at Netherne Hospital. Each sample of the categories studied comprised 20 subjects, 10 males and 10 females with age distributions as follows:

[1] Thanks are due to Drs. R. K. Freudenberg, physician superintendent, Netherne Hospital, for facilities given to carry out this research, and J. P. S. Robertson, senior psychologist there, for his encouragement and guidance.

Thanks are also extended to Drs. Doreen Firmin, physician superintendent, St. Lawrence's Hospital, and A. Walk, physician superintendent, Cane Hill Hospital, for additional cases to make up the cross-validating samples.

bright mental patients, mean 47.7, *SD* 9.47; dull mental patients, mean 45.7, *SD* 11.23; certified mental defectives, mean 34.2, *SD* 12.03; and brain-damaged patients, mean 50.2, *SD* 11.78. The first two categories were described psychiatrically as well-behaved chronic patients on parole. The allocation into bright or dull was made on the basis of Wechsler vocabulary raw scores, the limits of the bright category being 27.5 to the maximum, those of the dull category 18.5 to 22. None of the certified mental defectives had a Wechsler vocabulary raw score above 15. The neurological classification of the brain-damaged sample was as follows: paresis, five males and two females; cerebral atrophy, three males and four females; secondary cerebral carcinoma, one male; Huntington's chorea, four females; head injury, one male.

When the test results had been scored by the writer, they were independently scored by another person (not a psychologist) who was provided only with the actual results and the scoring instructions. There was only a very small proportion of disagreements; the product-moment correlation exceeded .90 for each test, using the difference method.

Results

Table 1 summarizes the results of the analyses of variance. It is clear that there are statistically significant differences among the categories in performance on each test. The factor of sex influenced only one test and that to a limited extent. The effect of the interaction was negligible.

In the case of the discriminant function analysis six equations were worked out: (*a*) bright versus dull; (*b*) bright versus mental defective; (*c*) bright versus brain-damaged; (*d*) dull versus mental defective; (*e*) dull versus brain-damaged; and (*f*) mental defective versus brain-damaged. These comparisons are set out in Table 2. The first column lists the comparisons as just enumerated. The second column gives the equations; here X = Marble-Board score, Y = Figure-Background score, and Z = Tactual-Motor score. The third column gives typical D for (A), the first member of a given comparison, and the fourth column typical D for (B), the second member of that comparison. The statistical significance of results is indicated as usual. In order to display the actual effectiveness of discrimination the cumulative percentages cut off by different values of D are shown in Table 3 for the four statistically significant comparisons and for the last comparison because of its special interest.

Table 1

Effects of Category and Sex: Analyses of Variance

Source	*df*	Mean squares: Marble Board	Figure Background	Tactual Motor
Category	3	131.950***	428.245***	44.292**
Sex	1	18.050*	103.512	0.013
Interaction	3	7.017	48.213	5.266
Error	72	4.580	51.820	10.587
Total	79			

* Significant at .05 level.
** Significant at .01 level.
*** Significant at .001 level.

Table 2

Comparison of Categories: Discriminant Function Analysis

Comparison	$D =$	Typical D for (A)	Typical D for (B)	F ratio ($df = 3$ by 36)
a	$0.90X + 0.11Y - 0.07Z$	9.00	7.36	2.28
b	$0.83X - 0.06Y + 1.28Z$	13.40	6.22	30.86***
c	$1.20X - 0.01Y + 0.15Z$	10.04	2.93	29.38***
d	$1.20X - 0.17Y - 0.03Z$	4.10	0.65	5.39**
e	$1.04X + 0.11Y - 0.13Z$	7.93	3.20	10.64***
f	$-0.15X + 0.25Y + 0.06Z$	4.43	2.72	2.46

** Significant at .01 level.
*** Significant at .001 level.

Table 3

Original Comparisons: Cumulative Percentages Cut Off

Value of *D*	Comparison *b*		*c*		*d*		*e*		*f*	
	B*	MD†	B	BD‡	D§	MD	D	BD	MD	BD
17 to 18	100									
15 to 16	80									
13 to 14	60	100	100							
11 to 12	45	95	90				100			
9 to 10	30	80	55	100			90			
7 to 8	15	65	25	95	100		55	100		
5 to 6	10	55	15	85	75		25	95	100	100
3 to 4		45		65	50	100	15	60	40	75
1 to 2		30		50	30	90		40	15	60
−1 to 0		5		30	10	25		20		20
−3 to −2		5		10	10	25				
−5 to −4					5	10				

* Bright sample.
† Mental defective sample.
‡ Brain-damaged sample.
§ Dull sample.

For the purpose of cross validation the three tests were repeated on fresh samples (partly drawn from adjacent hospitals) of 20 patients for each category chosen so as to correspond to the original samples in respect to age, sex, and Wechsler vocabulary level. The only divergence in neurological classification was that the case of head injury in the original sample has as its analogue in the second sample a case of tumor in the corpus callosum. As before, the test results were scored by the writer and independently by another person who was provided only with the actual results and scoring instructions; disagreement once again was trivial. The cumulative percentages cut off by different *D*

Table 4

Cross-Validating Comparisons: Cumulative Percentages Cut Off

Value of *D*	Comparison *b*		*c*		*d*		*e*		*f*	
	B*	MD†	B	BD‡	D§	MD	D	BD	MD	BD
17 to 18	100									
15 to 16	85									
13 to 14	70	100	100							
11 to 12	65	90	90							
9 to 10	55	90	65				100			
7 to 8	35	90	40	100			95	100		
5 to 6	10	80	15	90	100		45	90	100	100
3 to 4		65		70	95		10	65	50	95
1 to 2		35		50	55	100		45	25	50
−1 to 0		25		20	20	75		20	5	25
−3 to −2		15			5	35				
−5 to −4						20				

* Bright sample.
† Mental defective sample.
‡ Brain-damaged sample.
§ Dull sample.

values in the five relevant comparisons of the cross-validating samples are set out in Table 4. It will be noted that the correspondence between the original and the cross-validating samples is close.

Discussion

The fact that mental defectives are significantly discriminated from the bright and dull patients indicates the dependence of successful performance in these tests on general intellectual efficiency. The significant discrimination of the brain-damaged from the bright and dull patients without brain damage suggests that the Strauss-Lehtinen battery, if the Dot-Perseveration Test is dropped and a strict quantitative scoring system is employed for the other three tests, can be of assistance in this problem. The suggestion is strongly confirmed by the cumulative percentages data on the original and cross-validating samples. It is disappointing that there is no significant discrimination between mentally defective and brain-damaged patients, since no test has yet been devised which effectively answers this problem, and the statements of Strauss and Lehtinen (1) about endogenous and exogenous mental defectives have raised hopes that their battery might provide a solution. However, consideration of the cumulative percentages data on this comparison in Tables 3 and 4 indicates that at certain values of D (e.g., 1 to 2) a quite effective discrimination exists, which systematic item analysis might improve.

Summary

The Strauss-Lehtinen Battery for brain-injured children was applied to 80 adult mental hospital patients. The Marble-Board, Figure-Background, and Tactual-Motor tests were found, when quanitatively scored, to give jointly a successful discrimination between brain-damaged patients and bright or dull patients without brain damage. The discrimination between brain-damaged and mentally defective patients was much less satisfactory, but it is suggested that systematic item analysis might improve it. The findings were confirmed on cross-validating samples of 80 closely similar patients.

Received July 16, 1954.

Reference

1. Strauss, A. A., & Lehtinen, L. E. *Psychopathology and education of the brain-injured child.* New York: Grune & Stratton, 1947.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Normative Data on Progressive Matrices (1947)[1]

Meredith W. Green and Josephine C. Ewert

Mayo Clinic, Rochester, Minnesota

The Progressive Matrices (3) consist of 36 attractively colored problems in perceptual reasoning that, according to their author, can be presented in either booklet form or as a form board. At the beginning of the series the problems are easy to solve; they gradually become more complex, providing more intricate patterns and requiring more discrimination in making the correct choice.

According to the author, J. C. Raven, the Progressive Matrices, while not a test of general intelligence, will indicate a person's ability to form comparisons and to reason by analogy, as well as to show to what extent he is capable of organizing spatial perceptions into systematically related wholes. The test takes only about 20 minutes, thus affording an efficient means of gaining some information in those clinical situations in which a handicap of either motor coordination or of verbal ability is present.

The norms in Raven's manual are based on the scores obtained by 608 children between the ages of 5 and 11½ years in a rural community of a non-American culture. In our use of this test as part of the psychologic testing of children at the Mayo Clinic, it soon became evident that these norms appeared to be too low.[2] Therefore, we decided to carry out a normative study on children in our area to determine the expected score for various age levels and the relationship existing between the Progressive Matrices and the standard group intelligence tests used in the public schools in our region.

Method of Collecting Data

To facilitate the collection of a large number of test results quickly, the pages of the test booklet were reproduced on individual color slides, which permitted administration to an entire classroom at a time.

This method of administration of the matrices by slides obviously raised the question of the relationship between results of our group tests and results that would have occurred if the test had been given to these children individually. Data were obtained from children between the ages of 6–6 and 9–5; the 30 with the lowest scores on the slide form of the test were retested individually and 45 students were given the individual form prior to the group tests. The results of these variations in testing methods indicate a consistent tendency for the children to show a significant gain when retested, apparently as a result of practice. Since the gains were similar for each group administration (approximately 4 points), it was considered that, for the purpose of amassing normative data, the two methods of testing give comparable results. This agrees with Raven, who stated in his manual that there is apparently no real difference in test scores obtained from the booklet or form-board method of his test. The rank-order correlations were between .62 and .87 for these distributions, which are

[1] We are indebted to the Right Reverend Monsignor Louis D. O'Day, of St. John's Grade School, the Reverend R. J. Jansen, of St. Francis Grade School, Rochester, Minnesota, Superintendent H. H. Schaefer, of Stewartville, Minnesota, and Superintendent L. J. Gustafson, of Owatonna, Minnesota, for permitting this study to be conducted in their schools.

[2] In a letter to the authors, Raven explained that, while the sample of Dumfries school children was carefully compiled, and there is no reason to doubt its accuracy, he has the impression from repeated evidence collected by others that the intelligence of these Dumfries children is, on the average, lower than that for most of the urban districts of Britain. He also pointed out that there has been for several generations a steady movement of the population away from the rural districts into the cities or abroad.

sharply curtailed by age in both cases and by the narrow range of scores for those who had group tests prior to individually administered tests. This indicates that students tend to maintain their relative position within the group despite the effect of practice.

Our classroom testing included all the children in the second through sixth grades of four different elementary schools in the area of Rochester, Minnesota. The first two were parochial schools located in Rochester; they are estimated to have over 80 per cent of the local Catholic children. The third school was a consolidated one, located in a town of approximately 1,100 population; the fourth school included three-fifths of the children in a town of 10,000.

Results

The slide administration yielded scores for 1,214 children between the ages of 6–0 and 12–5 years. A comparison indicated no significant difference in scores for the children in any particular school. The product-moment correlation between age and score on the Progressive Matrices was .55 (σ_r, .02), indicating that in this group of children a positive and moderately high relationship existed between age and success on this test.

Table 1 shows by six-month intervals the number of children tested, the range of scores on Progressive Matrices, the mean score, the standard deviation, and the median. Also the Dumfries median score is quoted from Raven's manual. It is obvious that the median scores obtained on the 1,214 children from the Rochester area are consistently higher than those given for the Dumfries school children.

While Table 1 indicates that the median scores for children from the Rochester area show a steady increase that parallels increasing age, it is also evident that progression of scores is somewhat erratic. This irregular increase in median scores, together with the small amount of variation in median scores at times, seriously curtails confidence in the interpretation of individual scores at a definite age level. Both of these criticisms of the test, namely the need for more items so that chance errors would not distort the results and the ceiling effect at older ages, have been noted by other authors who have worked with the matrices.

Table 1

Scores for 1,214 Children on Progressive Matrices (1947)

Age (year/month)	Number of children	Mean score	*SD*	Range	Median	Dumfries median*
6/0–6/5	27	18.6	5.5	10–30	17	15
6/6–6/11	48	17.8	5.1	7–27	18	15
7/0–7/5	93	20.2	5.3	10–32	21	16
7/6–7/11	101	21.0	6.3	4–34	21	17
8/0–8/5	110	23.5	5.9	7–34	24	18
8/6–8/11	116	24.4	5.3	8–36	25	20
9/0–9/5	110	25.7	5.0	13–35	26	22
9/6–9/11	118	25.7	5.7	9–35	27	24
10/0–10/5	132	27.7	5.2	9–36	28	24
10/6–10/11	100	28.6	4.5	15–36	29	26
11/0–11/5	107	29.0	4.3	15–35	30	28
11/6–11/11	99	29.7	4.7	10–35	31	
12/0–12/5	53	29.1	5.2	14–35	31	

* Median for children at Dumfries, Scotland.

Our interest then shifted to what relationship existed between scores on Progressive Matrices and the results of standard intelligence tests. As would be expected, the relationship between scores on Progressive Matrices and Otis mental ages was high; this correlation was .78 for 192 children, all the students in one school between the ages of 6–6 and 12–5. Table 2 presents the correlation found between scores on the matrices and various intelligence tests (1, 2, 4). For various age levels, the correlations between IQ and scores on Progressive Matrices ranged from .19 to .69, but the concentration of correlation coefficients was somewhere near .40.

The frequency of success on each of the 36 items of the Progressive Matrices was computed; data from all children 7–6 to 12–5 years of age in the largest school were used. In Table 3 are recorded performances on the 12 most difficult items, the ones that elicited the smallest number of correct responses. The increase in the proportion of children responding correctly with increasing age was to be expected in view of the moderately high correlation (.55) already noted between age and scores on Progressive Matrices. The fact that these same 12 items were the most

Table 2

Relationships Between Scores on Progressive Matrices (1947) and Results of Otis, California Mental-Maturity, and Kuhlmann-Anderson Tests

Statistic	Age (year/month)					
	6/6–7/5	7/6–8/5	8/6–9/5	9/6–10/5	10/6–11/5	11/6–12/5
Number of children	22	31	51	49	47	36
Correlation with Otis IQ scores	.62**	.69**	.71**	.38**	.19*	.35*
Progressive Matrices, mean and *SD*	19.4± 5.2	20.0± 5.8	25.3± 5.8	27.1± 5.4	29.0± 3.6	30.6± 5.2
Otis IQ, mean and *SD*	113.2± 9.9	110.4± 7.0	114.2±10.5	110.1±10.8	109.3± 9.0	107.0± 8.5
Number of children	32	54	49	50	57	26
Correlation with California Mental-Maturity IQ	.42**	.37**	.42**	.47**	.43**	.45**
Progressive Matrices, mean and *SD*	22.7± 5.8	24.3± 6.5	25.9± 5.3	25.9± 6.0	28.5± 4.4	30.0± 3.6
California Mental-Maturity IQ, mean and *SD*	126.3±12.3	116.6±13.1	115.6±15.0	114.7±14.2	113.8±15.0	108.7±13.4
Number of children		64	72	83	59	61
Correlation with Kuhlmann-Anderson IQ		.28**	.56**	.48**	.52**	.37**
Progressive Matrices, mean and *SD*		22.7± 5.8	24.7± 5.1	28.2± 5.1	29.5± 5.0	29.0± 5.6
Kuhlmann-Anderson IQ, mean and *SD*		107.4± 8.9	101.5± 7.3	102.0±10.1	103.5±10.1	106.0±12.3

* Significant at .05 level.
** Significant at .01 level.

difficult throughout the age range of the children tested and the fact that there was a general tendency for an increase in correct responses to these items with increasing age argue for the consistency of the items.

Table 3

Percentage of Correct Answers on 12 Most Difficult Items of Progressive Matrices (1947) at Various Ages

Item	Age (year/month)				
	7/6–8/5 (*N**=75)	8/6–9/5 (*N*=75)	9/6–10/5 (*N*=90)	10/6–11/5 (*N*=63)	11/6–12/5 (*N*=63)
A 11	20	12	27	32	30
12	31	41	61	73	60
Ab 9	45	45	73	78	65
11	48	51	63	72	76
12	8	19	24	33	38
B 6	33	40	65	74	68
7	45	49	68	67	65
8	21	27	49	44	63
9	17	29	53	73	70
10	31	33	63	62	79
11	21	31	50	57	71
12	7	8	20	33	37

* *N* = number of children included.

Summary and Conclusions

The Progressive Matrices (1947), devised by J. C. Raven, were administered by means of slides to 1,214 children between the ages of 6–0 and 12–5 years in four elementary schools in Rochester, Minnesota, and nearby areas. Preliminary investigation indicated some effect from practice on repetition of the test but no significant difference between scores obtained by the slide form of administration and those obtained by the booklet form of the test.

The median scores for the children of the Rochester area were consistently higher than the median scores reported by Raven in the manual, with a range of differences from 2 to 6 points. Although the difference between the Rochester area median scores at adjacent age levels was frequently merely a point or so, the over-all correlation between scores on matrices and age was moderately high (correla-

tion of .55); a fairly substantial correlation was noted between the Otis mental ages and the scores on Progressive Matrices (correlation of .78). However, correlation with IQ's obtained from three different standard group intelligence tests used in these schools was only moderate (average correlation slightly greater than .40). Scores on Progressive Matrices correlated with the more verbal intelligence tests to about the same degree that they did with those tests which stress nonverbal aspects. Therefore, it would appear that the Progressive Matrices cannot be thought of as a test of nonverbal reasoning ability but instead should be considered as a test of fairly complex intellectual reasoning processes.

The frequency with which the items were answered correctly by a portion of the total group indicated that the items tend to maintain a stable order of difficulty; also, there was a gradual decrease in failure on items over the age range from 6–0 to 12–5 years.

Received July 12, 1954.

References

1. Kuhlmann, F., & Anderson, Rose G. *Kuhlmann-Anderson Intelligence Tests.* Minneapolis: Educational Test Bureau, 1952.
2. Otis, A. S. *Otis Quick Scoring Mental Ability Tests.* Yonkers, N. Y.: World Book Co., 1952.
3. Raven, J. C. *Progressive Matrices (1947).* London: H. K. Lewis, 1947.
4. Sullivan, E. T., Clark, W. W., & Tiegs, E. W. *California Test of Mental Maturity.* Los Angeles: California Test Bureau, 1951.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Relationships Between Verbal and Nonverbal Parts of the CTMM and WISC

Grace T. Altus
Santa Barbara County Schools

One of the reasons for the increasing preference for the WISC (3) over the Stanford-Binet is undoubtedly its separation into Verbal and Performance scales. Similarly, choice of the California Test of Mental Maturity, or CTMM (2), as a group measure of intelligence is frequently made because it yields both Language and Non-Language IQ's. The tacit assumption is often made, as by school psychologists using both tests, that the corresponding verbal and nonverbal portions of the two tests are highly correlated. Some data are here presented bearing on the validity of this assumption.

The subjects were 100 elementary school children, 71 boys and 29 girls, referred to the Guidance Department of the Santa Barbara County Schools by their teachers. The modal reason for referral was consideration for classes for the mentally retarded ($N = 36$), while the remaining cases were referred for a variety of reasons including grade placement, academic disabilities, emotional disturbances, and the like. All children were given the WISC by the writer and the Elementary Short Form of the CTMM in their regular classrooms as part of a routine county testing program. In the cases of 66 children the WISC and CTMM were given less than a year apart, while in no case were the tests as much as three years apart.

Table 1 describes the composition of the group with respect to obtained IQ's on the two tests. It can be seen—as might be expected from the method by which subjects were selected—that the sample is on the average less intelligent but also more variable than an unselected population. The similarity of means on corresponding parts of the two tests is noticeable.

Although the wide sigmas for both tests would tend to raise intercorrelations spuriously, there were other factors operating to reduce such r's: the fact that the group tests were given and scored by teachers in an informal school setting rather than under precise experimental conditions, the time intervening between tests, the inclusion of children with serious reading problems who would be handicapped on the group test, and the like.

Nearly half the children ($N = 49$) had mixed language backgrounds, primarily Spanish-English. Somewhat surprisingly, the majority of this bilingual sample did better on the group test than on the individual one on both verbal and nonverbal sections, with the trend reversed for the unilingual sample. Within both the WISC and the CTMM, however, the bilingual group did conspicuously better in the nonverbal portions than on the verbal portions.

Table 1

Means and Standard Deviations of WISC and CTMM Intelligence Quotients

Test	Mean	*SD*
WISC		
Verbal	82.7	17.2
Performance	89.8	16.3
Full Scale	84.5	17.4
CTMM		
Language	81.8	17.3
Non-Language	88.9	20.0
Total	84.8	17.4

Table 2

Intercorrelations of WISC and CTMM Intelligence Quotients

	WISC		CTMM		
Test	Perf.	Full	Lang.	Non-Lang.	Total
WISC					
Verbal	.77	.93	.71	.65	.76
Performance		.92	.57	.67	.68
Full Scale			.70	.68	.77
CTMM					
Language				.57	.89
Non-Language					.85

Thus, 76% of the bilingual group had higher IQ's on the CTMM Non-Language than Language test, with 27% of the group exceeding an IQ discrepancy of 20 points in this expected direction. These findings would appear to conform to what one might theoretically expect for the test pattern of a bilingual—and in this case also economically underprivileged—group: that is, evidence of greater ability on tests not requiring verbalization. Similar findings for the WISC were reported by the author elsewhere (1).

Table 2 shows the interrelationships derived from Pearsonian correlations of the individual and group tests and their subportions for the entire sample of 100 children. Whereas the corresponding verbal portions of the two tests are substantially correlated, and the corresponding nonverbal portions only slightly less so, it would appear that the CTMM Non-Language has considerably less overlap with its partner scale, the CTMM Language, than the WISC Performance has with the WISC Verbal. The analyses of the two tests, then, cannot be considered fully parallel. In the absence of external validation of the two nonverbal tests, one cannot be sure what the difference represents in behavioral correlates or aptitudes, but surely it is a difference.

Further analysis of test scores was made by tabulating IQ differences for individual children on the three pairs of corresponding IQ's—verbal, nonverbal, and total—of the WISC and CTMM. The nonverbal tests showed the greatest number of wide discrepancies, with one child in four showing a difference in excess of 18 IQ points between WISC Performance and CTMM Non-Language IQ. The difference between *total* IQ's on the two tests was, however, less than 10 IQ points in 65% of the cases, indicating reasonably adequate individual prediction from one test to another.

In short, it would seem justified to conclude that, within a comparable school referral setting, the WISC and CTMM are markedly comparable as to group assessment and roughly comparable as to individual scores and major breakdown into verbal and nonverbal abilities. The precise differentiation of the latter abilities deserves further exploration, could studies be devised to determine the actual meaning of nonverbal as opposed to verbal tests. The fact that bilingual children did better on the respective nonverbal tests than on the verbal appears to be evidence of their intrinsic validity. In any event, the saturation of verbal factors in virtually all the tests herein discussed would imply their usefulness in a school setting.

Received August 26, 1954.

References

1. Altus, Grace. WISC patterns of a selective sample of bilingual school children. *J. genet. Psychol.*, 1953, **83**, 241–248.
2. Clark, W. W., *et al. California Short-Form Test of Mental Maturity; manual.* (Elementary 1950 S-Form.) Los Angeles: California Test Bureau, 1950.
3. Wechsler, D. *Wechsler Intelligence Scale for Children; manual.* New York: Psychological Corp., 1949.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Food Aversions: Some Additional Personality Correlates

Wendell I. Smith, Elizabeth K. Powell
Bucknell University
and Sherman Ross
University of Maryland

One of the interesting aspects of human behavior that has been relatively unexplored is the area of food habits. There are many facets of research which need to be undertaken to assay the variables that enter into this complex universal behavior (8). Foods are a readily available portion of the behavioral environment at which responses can be directed. A series of studies which deal with food aversions has been reported. In some of these (3, 11, 12) neurotic *Ss* have been found to select a larger number of foods which they "dislike and would not eat" than have normal controls.

We recently reported on the relationship between scores on the Taylor Manifest Anxiety Scale and food aversions in a college student population (10). A low positive (but significant) correlation was found between scores on the *A* scale and the number of food aversions. Sampson and Bindra (9) have questioned the meaningfulness of high scores on the *A* scale when used with neurotic *Ss*. They find that scores in the second and third quartiles were more predictive of neuroticism than scores in the top quartile. These findings may explain the low degree of relationship which we found.

The present study stems from our continued interest in the relationships between food aversions and personality characteristics. We hope to determine the relationship between some additional personality factors and performance on a check list of foods.

If a large number of food aversions is related to neurotic trends, if the overriding characteristic of the neurotic is fear and anxiety, and if the fear and anxiety can be assumed to displace to several areas of behavior, then we can predict that (*a*) *Ss* having a large number of food aversions will have marked neurotic characteristics, and (*b*) such *Ss* will tend to report that they dislike and will not eat foods which they have never tasted. The second prediction could represent a socially acceptable way of expressing *Ss*' fear and anxiety.[1] More simply, those *Ss* who have many aversions for foods which they have tasted will generalize this aversion to foods which they have never tasted.

Method

A check list of 28 selected foods (see Table 2) was given to 235 undergraduate students (117 male and 118 female) in introductory psychology classes. The following instructions accompanied the check list:

> Indicate which of the foods in the list below you refuse to eat. Do this as follows: Read over the list slowly. When you come to the name of a food you dislike, put a circle around the number of that food. After marking all the foods you do not like, go back over the list and draw a circle around the *names* of all the foods which you dislike, but have never tasted. After each food you dislike, state the reason for your dislike.

Parts of three paper-and-pencil inventories which purport to evaluate characteristics which we believed might be related to the number of food aversions were also adminis-

[1] We are indebted to Dr. J. W. Gustad, University of Maryland, for assisting us in developing this approach.

tered to this group. The inventories used were: P-S Experience Blank, Part 1 Physiological (5); The Personal Audit, Form LL, Part V (Stability-Instability) and Part VII (Emotionality-Steadiness) (1); and the F scale of authoritarianism (2). Additional information concerning *Ss*' sex and age, father's occupation, religious preference, frequency of attending church, etc., was requested. Names or other identification was not required.

Results and Discussion

Two food aversion scores were computed for each *S*. The first (FA-I) was the total number of foods which *S* checked as disliked. The second score (FA-II) was FA-I minus the number of foods checked as disliked but which *S* had never tasted. FA-II then is the number of foods which *S* has tasted and for which he has an aversion. The results obtained from this procedure should be fairly close if *Ss* have very few aversions for foods which they have never tasted. The relative magnitude of the difference between FA-I and FA-II should provide information on whether or not familiarity with a food is an important determinant of the number of food aversions (3).

From the total sample of *Ss* we drew two samples: High FA and Low FA. The cutoff for the High group was 6 or more checked items, while the cutoff for the Low group was 1 or 0 items. Two such groups were made for each of the FA scores as follows: (*a*) FA-I High ($N = 51$, Mean $= 7.2 \pm 1.4$) and FA-I Low ($N = 51$, Mean $= 0.6 \pm .5$); (*b*) FA-II High ($N = 31$, Mean $= 9.8 \pm 2.3$) and FA-II Low ($N = 40$, Mean $= 0.6 \pm .5$). The mean score for each of the groups on each inventory was computed and tests of the significance of the difference between the High and Low FA groups on each test were made. These results are presented in Table 1.

It is clear from Table 1 that the High and Low groups differed significantly on all but the F scale. In all cases, the mean scores for the High FA groups tended more toward those obtained for neurotic populations than did the mean scores for the Low FA groups.

Interestingly enough, the results on the authoritarian scale for both the High and Low FA groups were nonsignificant even though the description of the authoritarian personality would lead one to expect that the F scale might differentiate between those *Ss* who are characterized by relatively high anxiety, relatively poor emotional control, instability, feelings of inadequacy, and physiological dysfunction, and those who are normal in these characteristics. Masling (7) has made this point previously. That the F scale scores of very extreme groups (extreme in that they reported not more than two aversions or nine or more aversions) may differ significantly is indicated by the results which we obtained from a comparison of the 15 highest- and 15 lowest-scoring *Ss*. These *Ss* were all females. A similar analysis for 30 extreme male *Ss* did not result in a significant difference for F-scale scores. This adds to the evidence to be presented that there are marked sex differences in the number of food aversions.

In our earlier study we proposed that the data obtained from a check list of foods might be more meaningful if information on experi-

Table 1

Personality Test Results for High and Low FA Groups

	FA-I				FA-II			
	High		Low		High		Low	
Score	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Emotionality	56.6*	11.9	33.3	11.6	35.9*	11.7	29.9	12.2
Stability	41.2*	22.4	31.0	18.7	45.4*	21.6	30.7	20.1
P-S inventory	119.7*	42.3	90.2	45.0	118.0*	43.6	79.1	51.9
F scale	−7.9	21.5	−5.7	22.4	−9.0	21.0	−4.5	23.9

* Difference between High and Low groups significant at the .01 level of confidence.

ence with each food and the reasons for disliking each food were available. We made an attempt in this study to collect these data by (*a*) having each *S* signify which foods he had never tasted and (*b*) having each *S* state the reason for his aversion to each food checked. From our analysis, it appears that the two food aversion scores have approximately the same discriminative power. Low and High groups whether selected by FA-I scores or by FA-II scores were significantly different on three of the four inventories. This finding was derived from analysis of the data without regard to sex. As we will now show, when the analysis is made taking sex into account, the two food aversion scores for females are significantly different.

Comparisons were made between male ($N = 117$) and female ($N = 118$) *Ss* on each of the inventories as well as on food aversions. Only on the trait of stability-instability was the difference between the sexes significant. The female *Ss* as a group were somewhat less stable (CR = 3.78). As in previous studies, female *Ss* reported more food aversions than did male *Ss*. For FA-I, the mean for male *Ss* was 3.1 ± 1.7 aversions and for female *Ss* it was 5.3 ± 2.9 aversions. For FA-II the mean for male *Ss* was 2.6 ± 2.7 aversions and for female *Ss* 4.3 ± 2.6 aversions (CR = 2.22). On both scores female *Ss* were significantly higher than males. Further, the FA-I mean for females is significantly higher (CR = 2.89) than the FA-II mean, while for males there is no significant difference between the mean FA scores. Seemingly, for our sample of females there was a tendency to dislike foods which had never been tasted, indicating generalization.

A further check for this factor was made by comparing those *Ss* who reported aversions to foods they had never tasted with those *Ss* who reported that they had tasted each of the foods for which they had an aversion. This was done for female *Ss* only. Fifty-six of the females reported one or more aversions to foods never tasted. The mean FA-I score for this group ("generalizers") was 6.5 ± 1.7. A group of 25 *Ss* was drawn from the remainder of the female sample on the basis of their FA-I scores. The mean for this group ("nongeneralizers") was 6.1 ± 1.4.

Table 2

Percentages of Male and Female *Ss* ($N = 235$) Reporting Food Aversions for Each Food Item

Food	Male	Female
Mushrooms	16	26
Cottage cheese	22	19
Swiss cheese	11	17
Tea	6	21**
Bean soup	6	21**
Salmon	9	13
Lima beans	11	14
Radishes	6	11
Cabbage	8	14
Cantaloupe	3	4
Potato soup	9	16
Grapefruit juice	3	9*
Veal chops	2	2
Fried eggs	3	14**
Tomatoes	0	3*
Cherries	0	6**
Pears	0	3*
Salt pork	9	11
Liver	21	37**
Kidneys	18	25
Brains	14	14
Gravy	0	3*
Chili	4	15**
Chocolate	1	1
Beer	12	32**
Buttermilk	45	49
Onions	8	13
Mush	13	19

* Sex difference significant at the .05 level of confidence.
** Sex difference significant at the .01 level of confidence.

Thus, the total number of aversions was held constant and a chi-square test was made to determine whether or not the two groups differed in their means for each of the personality measures. A chi square of 9.575 (3 *df*, $p < .02$) was obtained. The groups differed significantly only on the F scale, on which the nongeneralizers were found to be much less democratic than the generalizers. When the generalizers are compared with the full sample of female *Ss*, the mean FA-I score and all inventory scores are practically identical. These results indicate that, although more female *Ss* than male *Ss* report aversions for foods which have never been tasted, the female *Ss* so reporting do not differ in test performance from the total female sample. No ready explanation occurs to us for the lower F-scale score of those *Ss* who did not report aversions to foods never tasted.

In Table 2 the sex differences in aversions for each of the 28 foods are presented. Significantly more female *Ss* than male *Ss* disliked 11 of the foods. When we compare the present findings with those obtained in our previous study, the differences between the sexes overlap on only two foods, chili and beer, indicating that the two samples differed in the frequency with which each food was disliked.

In order to clarify the differences between the extremes of our sample, the 15 *Ss* having the highest FA-II scores and the 15 *Ss* having the lowest FA-II scores were selected for comparison. Since the High *Ss* were all females, only females were selected for the Low group. The Low group mean was 1.3 aversions, while the High group mean was 10.8 aversions. The mean inventory scores for the Low group were well within the range for "normals" while the scores of the High group bordered on those typical for neurotics. All the differences between the two groups were statistically significant as determined by chi-square tests comparing obtained scores on each inventory with expected scores on each inventory. The chi-square values (with 1 *df*) obtained from comparing these Low and High female groups on each test were: P-S Experience Blank, 80.5 ($p < .001$); Stability-Instability, 15.6 ($p < .001$); Emotionality-Steadiness, 14.7 ($p < .001$); and F scale, 4.12 ($p < .05$).

Examination of the biographical data for these 30 *Ss* did not reveal any apparent differences between the groups in age, parents' education, religious preference, father's occupation, size of home community, and regularity of attending church. There were two and one-half times as many only children in the High group but the sample is obviously too small to attach great significance to this factor.

When *Ss*' reasons for their aversions are considered, one is struck with the superficiality of the answers. After examining the responses, we decided against any formal categorization. Slightly less than half of all answers consisted of "I don't like the taste." This response leaves us little more information than we had when we started. A comparison of the reasons stated by the Low group and by the High group on FA-II scores provides only one interesting point. Forty per cent of the reasons given by the High group could be categorized as responses which would probably not usually characterize the food. For example, such adjectives as "mushy" and "slimy" were applied to foods such as liver, tomatoes, beans, and chili. Such descriptions as "it gives me indigestion," "I don't like the feel of it," and "I can't even stand the thought of it" were common. These statements were applied to such commonly consumed items as beans, tomatoes, liver, cabbage, cherries, eggs, onions, and beer. This point is of interest, since only one response in the Low group could be so categorized. This single response was given to fried eggs. The *S* wrote: "I am allergic to concentrates of protein." It is unfortunate that our *Ss* were anonymous, since it might have been valuable to interview them intensively concerning the reasons for their aversions.

The results of this study give indirect support to the hypothesis that there is some generalization of affect to the socially accepted area of food habits. There is a tendency for individuals who report disturbances in affective behavior to report an unusual number of food aversions.

Lewin's (4) contention that "people like what they eat, rather than eat what they like" seems to be an insightful inference. The *Ss* in our study did not like those foods with which they had had no experience. Familiarity of the food appears to be an important factor.

The bases for the aversions reported still remain to be clarified. Intensive interviewing might result in fresh leads to this problem. An interesting lead to be explored in future studies is supplied in Margaret Mead's statement that food habits may be influenced by: "The Puritan tradition of a connection between food which is disliked, and the tendency in communities with a Puritan tradition to use food for purposes of reward and punishment . . ." (7, p. 23).

Summary

This study deals with the relationships between food aversions (FA) and personality characteristics. A group of 235 undergraduate students was studied for possible relationships between FA and scores made on four selected personality inventories. These inventories were: (*a*) P-S Experience Blank, (*b*) The Personal Audit (Stability-Instability), (*c*) The Personal Audit (Emotionality-Steadiness), and (*d*) F scale of authoritarianism.

Two kinds of FA scores were computed and used to categorize the sample in a variety of ways for comparison purposes. The first score was based on the total number of items checked as disliked from the list of 28 foods given. The second score was based on the number of items checked as disliked but which *S* reported he had tasted. The basic comparisons were made between selected High and Low groups for each of these scores, and for sex differences using the total sample.

The results permit the following conclusions:

1. High and Low FA groups on each scoring technique differed significantly on all of the personality inventories, except for the F scale. The High groups were closer to the scores obtained by neurotic *S*s on these inventories.

2. Female *S*s ($N = 118$) reported significantly higher FA scores than male *S*s ($N = 117$).

3. No significant differences were found between the two FA scores of the male *S*s. A significant difference was found between the two FA scores for female *S*s. This result was interpreted to mean that there was a tendency for female *S*s to dislike foods which they had never tasted.

4. The stated reasons for an FA by *S*s were most frequently in the form of descriptions which would probably not be the usual characterization of the food.

Some limited interpretations were made of these findings in regard to the role of foods as socially acceptable objects against which the effects of anxiety and fear can be directed.

Received July 6, 1954.

References

1. Adams, C. R., & Lepley, W. M. *The Personal Audit* (Forms L and S). Chicago: Science Research Associates, 1941.
2. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.
3. Gough, H. G. An additional study of food aversions. *J. abnorm. soc. Psychol.*, 1946, **41**, 86–88.
4. Lewin, K. Forces behind food habits and methods of change. In C. E. Guthe [Chm.], The problem of changing food habits; report of the Committee on Food Habits. *Nat. Res. Coun. Bull.*, 1943, No. 108. Pp. 35–65.
5. McFarland, R. A., & Seitz, D. P. A psychosomatic inventory. *J. appl. Psychol.*, 1938, **22**, 327–339.
6. Masling, J. M. How neurotic is the authoritarian? *J. abnorm. soc. Psychol.*, 1954, **49**, 316–318.
7. Mead, Margaret. In C. E. Guthe [Chm.], The problem of changing food habits; report of the Committee on Food Habits. *Nat. Res. Coun. Bull.*, 1943, No. 108. Pp. 20–31.
8. National Research Council. Committee on Food Habits. Manual for the study of food habits. *Nat. Res. Coun. Bull.*, 1945, No. 111.
9. Sampson, H., & Bindra, D. "Manifest" anxiety, neurotic anxiety, and the rate of conditioning. *J. abnorm. soc. Psychol.*, 1954, **49**, 256–259.
10. Smith, W. I., Powell, Elizabeth, & Ross, S. Manifest anxiety and food aversions. *J. abnorm. soc. Psychol.*, 1955, **50**, 101–104.
11. Wallen, R. Food aversions of normal and neurotic males. *J. abnorm. soc. Psychol.*, 1945, **40**, 77–81.
12. Wallen, R. Food aversions in behavior disorders. *J. consult. Psychol.*, 1948, **12**, 310–312.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

The Effect of Pinhole Observation on the Flicker Fusion Frequency in Patients with Brain Pathology[1]

James C. Stauffacher, John B. Marks
VA Hospital, American Lake, Washington

and Albert F. Ax
University of Washington

This study was designed to determine whether or not the lowered critical flicker fusion frequency (CFF) of patients with brain pathology is a result of alterations in the size of the pupil and whether use of a restricted pinhole pupil might not increase the effectiveness of CFF examinations in screening patients with brain pathology.

The experimental group consisted of 21 patients with various organic brain disorders. The control group consisted of 21 psychotic patients with no known brain pathology. Each patient of the brain pathology group was matched in age with a control patient.

The intermittent visual stimuli were obtained by using a Strobotac to activate a neon bulb which was painted black except for an orifice 3/8 inch in diameter, through which the glow plate was viewed. The bulb was mounted at one end of a lightproof box 40 inches from the viewing cowl from a stereoscope. One eye could be shielded by the insertion of a plate, and a second plate with a pinhole of .046-inch diameter could also be inserted before the other eye.

Each patient was tested for CFF under three conditions: with the usual binocular stimulation, monocular stimulation, and monocular stimulation through the pinhole. The order of these conditions was alternated systematically and the establishment of a "time set" was avoided by systematic alteration of the beginning frequency of the light stimulus. Fatigue was minimized by having subjects close their eyes between observations.

The usual binocular condition discriminated between the brain pathology and control groups at the .01 level of confidence when the mean of the up trials (28.05 and 31.18), the mean of the down trials (27.04 and 31.30), and the mean of the up and down trials (27.55 and 31.25) were considered. When compared with binocular observation, the monocular condition produced slightly smaller differences between the pathology and control groups and the difference between the groups reaches the .01 level of confidence for each criterion used. When pinhole observation is used, the pathology group has lower CFF in mean up trials (17.54 and 18.20), mean down trials (15.94 and 17.89), and mean of up and down trials (16.75 and 18.04). However, the differences are smaller and the reliability is greatly reduced, the significance levels being .25, .01, and .05 respectively.

These results indicate that differences in CFF between pathology and normal groups cannot be explained in terms of variations in pupil size. The usual binocular observation yielded a more reliable difference between pathology and control groups than did the pinhole observation used. The factors responsible for the reduction in the reliability of pinhole observation are not indicated from the study.

Brief Report
Received November 23, 1954.

[1] An extended report of this study may be obtained without charge from J. C. Stauffacher, VA Hospital, American Lake, Wash., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4467 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 19, No. 2, 1955

Emotional Adjustment and Dieting Performance[1]

John Summerskill and C. Douglas Darling
Department of Clinical and Preventive Medicine, Cornell University

Two recurring observations on obesity suggest its relationship to psychosomatic disorders. First, body weight is directly related to caloric intake (3, 4, 5, 9). Secondly, caloric intake is in some degree and in some manner related to the individual's emotional state (1, 2, 3, 6, 7, 8, 10). Consequently, the treatment of obesity necessarily involves dieting, but treatment is successful only if emotional conditions are favorable. This research was focused on the latter relationship: emotional influences on dieting performance.

The immediate problem was posed in experiments conducted at Cornell University on the efficacy of certain diets (11, 12). Subjects were given three carefully measured meals daily for three months. Evaluation of dietary effects involved continuing weight checks, basal metabolism studies, fecal balance studies, blood sample analyses, etc. The validity of the findings from these intensive investigations was largely dependent upon the ability of the subjects to adhere rigorously to the diet regime for a full three-month period. From this standpoint it was important to select subjects who were able to diet successfully.

The authors hypothesized that there is a positive relationship, subject to quantification, between emotional stability and dieting success. Validation of a technique for selecting dietary research subjects was undertaken on the assumption that it would be relevant to the diagnosis and treatment of obese patients in clinical settings. Suitability for dieting is important in view of the recognized possibility that acute disturbance can be precipitated in maladjusted individuals when eating habits are radically altered (2).

[1] The diet experiments, conducted by the School of Nutrition, Cornell University, were supported in part by the Gannett Nutrition Fund and in part by an appropriation from the State of New York through the State University. The authors wish to express their appreciation to Norman S. Moore, M.D., Jeffrey Fryer, M.D., and Charlotte M. Young, Ph.D., of the School of Nutrition, for their interest and cooperation in the psychological phase of the research.

Method

Psychological tests were administered to subjects of three dietary experiments conducted in the spring of 1952, 1953, and 1954. The total number of subjects tested was 29, including 17 college women and 12 college men with an average weight excess of 25%. Scores on emotional adjustment were subsequently compared with: (*a*) pounds lost through dieting and (*b*) diet performance ratings. The second criterion was included in the design because pounds lost does not adequately represent dieting success as it relates to individual differences in percentage of overweight, activity expenditure, and other factors. Each dieter was rated on his demonstrated acceptability (+ or −) for controlled dieting by the physician, nutritionist, or dietician who had supervised the dieting. Two independent ratings were available for 23 subjects and these were in agreement in all cases but one. Specific methods employed in each of the three experiments were as follows:

Experiment I (1952). During the course of the first dietary experiment the nutritionist learned that not all subjects were adhering strictly to the diet provided. At the completion of dieting, this situation was brought to the attention of the writers who then studied the six of ten subjects still on campus. Each subject was seen in an evaluative psychiatric interview and was administered a battery of psychological tests including: (*a*) Bell Ad-

justment Inventory, (*b*) Bernreuter Personality Inventory, (*c*) MMPI, and (*d*) Thurstone Personality Schedule.

Experiment II (1953). The results of Experiment I suggested that prediction of dieting performance was feasible from a single, objective index of emotional adjustment. There were also indications that adjustment during dieting was largely expressed through psychosomatic complaints. Accordingly, the 11 subjects of the second dietary experiment were tested before dieting began with the: (*a*) Bell Adjustment Inventory and (*b*) Psychosomatic Inventory. These tests were repeated at the midway point and at conclusion of the diet.

Experiment III (1954). To cross-validate the selection procedure and increase *N* the two tests employed in Experiment II were administered to 26 applicants for the third dietary experiment. Twelve subjects were then selected for the diet, each above average on test measures of emotional stability.

Results

A positive relationship between above average emotional adjustment and dieting success was found within each experimental group and for the experimental groups combined. Table 1 shows the dieting performance of subjects who demonstrated better than average vs. average or below average emotional adjustment according to the college norms of the Bell Adjustment Inventory. Considering all experimental groups, 78% of the well-adjusted individuals were independently rated as acceptable dieters whereas only 18% of the other individuals were so rated. This relationship is highly significant ($p > .01$) according to a chi-square test on the fourfold table (Yates' correction for small frequencies employed). With regard to the other criterion, well-adjusted individuals lost an average of 21.0 lb. ($SD = 5.25$) whereas other individuals lost an average of 13.6 lb. ($SD = 5.17$). A t test shows that this difference is highly significant ($p > .01$).

Table 1 contains additional data on the predictive value of the emotional scale of the Bell Adjustment Inventory. Dieters selected for Experiments I and II without the use of test scores proved to be acceptable in 7 of 17 cases. On the other hand, dieters selected for Experiment III on the basis of test scores proved acceptable in 9 of 12 cases. In the three groups test scores accurately classified dieters as acceptable or unacceptable in 23 of 29 cases.

Other observations, some based on numbers too small for statistical treatment, were as follows:

1. Psychological adjustment scores from the P-S Inventory (23 cases) were in essential agreement with Bell emotional scores and showed a similar positive relationship with dieting performance.
2. Psychiatric interview findings on emotional stability (6 cases) were in essential agreement with test indices of emotional adjustment.
3. Indices of adjustment in other areas (e.g., the Bell home, health, and social adjustment scales—29 cases) were not related to dieting performance.
4. Test scores before, during, and after dieting (11 subjects) were essentially unchanged, with one exception. The number of somatic symptoms reported was somewhat greater during dieting.

Table 1

Diet Performance Ratings and Mean Weight Loss of Dieters with Above Average and Average or Below Average Bell Emotional Adjustment Scores

Experimental group	Dieters with above average Bell emotional adjustment scores			Dieters with average or below average Bell emotional adjustment scores		
	N	Number rated acceptable	Mean pounds lost	*N*	Number rated acceptable	Mean pounds lost
Experiment I (1952)	2	2	22.5	4	1	14.4
Experiment II (1953)	4	3	17.5	7	1	13.1
Experiment III (1954)	12	9	21.9	0		
All groups	18	14	21.0	11	2	13.6

5. Analysis of Bell emotional adjustment items showed no consistent pattern of symptoms among maladjusted dieters (11 cases). Similarly, there was no single or unique personality pattern evident in the psychiatric interviews of three dieters of inferior emotional stability.

Discussion

It is recognized that results are based on a small sample of college students and that extension of conclusions to a general population will require further research. Confidence in the demonstrated relationship between tested emotional adjustment and dieting performance in college students is strengthened by the consistency of findings in three discrete experiments. With this in mind, the writers feel that objective tests of emotional adjustment, e.g., the Bell or P-S Inventory, can be profitably used in selecting individuals who can handle the pressures involved in controlled dieting.

The findings also support the authors' clinical observation that serious emotional disturbance may or may not be present with obesity. Psychiatric case studies of obese patients sometimes leave the impression that obesity invariably reflects serious deep-seated conflict. The present results indicate that obesity can be associated with varying states of emotional adjustment from the essentially normal to the seriously disturbed. When signs of good emotional adjustment are present, the chances that dieting will be successful appear to be about four times as great as they are when these signs are absent.

Summary

The purpose of the research was to examine the relationship between emotional adjustment and dieting performance in order to devise a procedure for selecting individuals suitable for controlled dieting. A statistically significant relationship was found between objective test measures of emotional adjustment and (*a*) pounds lost through dieting and (*b*) independent diet performance ratings. The results show that psychological instruments such as the Bell Adjustment Inventory, or Psychosomatic Inventory, can be profitably used in eliminating individuals whose participation in rigorous dieting can have undesirable consequences. The findings also indicate that obesity can be associated with varying states of emotional adjustment from the essentially normal to the seriously disturbed.

Received June 24, 1954.

References

1. Bruch, Hilde. Psychological aspects of obesity. *Bull. N. Y. Acad. Med.*, 1948, **24**, 73–86.
2. Bruch, Hilde. Psychological aspects of reducing. *Psychosom. Med.*, 1952, **14**, 337–346.
3. Hamburger, W. W. Emotional aspects of obesity. *Med. Clin. N. Amer.*, 1951, **35**, 483–499.
4. Newburgh, L. H. The cause of obesity. *J. Amer. med. Ass.*, 1931, **97**, 1659.
5. Newburgh, L. H. Obesity. *Arch. int. Med.*, 1942, **70**, 1033.
6. Reeve, G. H. Psychological factors in obesity. *Amer. J. Orthopsychiat.*, 1942, **12**, 674–679.
7. Rennie, T. A. C. Obesity as a manifestation of a personality disturbance. *Dis. nerv. Syst.*, 1940, **1**, 238–247.
8. Richardson, H. B. Obesity and neurosis. *Psychiat. Quart.*, 1946, **20**, 400.
9. Rynearson, E. H., & Gastineau, C. F. *Obesity*. Springfield, Ill.: Charles C Thomas, 1949.
10. Schick, A. Psychosomatic aspects of obesity. *Psychoanal. Rev.*, 1947, **34**, 173–183.
11. Young, Charlotte M. Weight reduction using a moderate-fat diet. *J. Amer. diet. Ass.*, 1952, **28**, 410–416.
12. Young, Charlotte M., Ringler, I., & Greer, Betty J. Reducing and post-reducing maintenance on the moderate-fat diet: metabolic studies. *J. Amer. diet. Ass.*, 1953, **29**, 890–896.

NEW BOOKS AND TESTS

Books

Bauman, Mary K. (Ed.) *Adjustment to blindness.* Harrisburg, Pa.: Commonwealth of Pennsylvania, State Council for the Blind, 1954. Pp. xiii + 198.

This intensive study of 443 blind persons from six states is a real contribution to the literature of adjustment to handicaps. The three groups studied were: (*A*) employed and generally well-adjusted ($N = 162$), (*B*) well-adjusted except in employment ($N = 150$), and (*C*) generally poorly adjusted, including employment ($N = 131$). They were first found not to differ in sex, color, residence, birthplace, height, weight, or age. (Group *B* was made up of somewhat younger people.) With these factors constant, differences appeared in favor of well-adjusted employed blind in intelligence and personality inventory scores. These factors overshadowed amount of visual loss, health, education, and family and social interaction as influential causes. It would have been more satisfying to the critic to have a clearer discussion of the use of the criteria in setting up and separating the three groups. Although most of the criteria were objectified and tabulated, their importance was minimized by reporting them along with the dependent variable. Since a personality inventory adapted to the blind was administered rather than a standard one, it would have been well to have included it, in addition to a description of its parts and reliability coefficients. This report presents a prodigious amount of reliable information on employment, recreation, home life, travel, and health which will greatly assist those interested in the rehabilitation of the blind. A helpful 123-title bibliography concludes the volume.—*F. McK.*

Holsopple, James Quinter, & Miale, Florence R. *Sentence completion.* Springfield, Ill.: Charles C Thomas, 1954. Pp. xiii + 177. $5.50.

This little book has many charms. First, it is packed with illustrative protocols of sentence completions by persons described in sufficiently relevant detail for the reader either to check his own interpretations, or to determine for himself whether or not the responses really have the flavor of genuine clinical utility. Second, it contains an extensive case interpretation that is a model of teaching by example. Third, the book is thoroughly free from both clinical jargon and extravagant claims. Holsopple and Miale overtly deny that the sentence completion method is a "test," constructed on the model of a yardstick or a clock, and argue that the task involved in using sentence completions is that of sensitive interpretation, not of scoring. This appropriate frankness leads them to emphasize the point that the clinician is actually part of the instrument in the sentence completion technique, as in virtually all projective devices. It leads them away from any attempt to quantify, beyond a voluminous listing of the most frequent completions to their particular stems. As a result, they implicitly raise the two central issues of what attributes characterize a competent clinician and to what extent clinical psychology may accept Thorndike's dictum about the measurability of all things. That Holsopple and Miale do not deal directly with these problems is hardly to their discredit. They have faced them in a context that is instructive, reflective of an intimate and sympathetic knowledge of the clinician's job, and fruitfully suggestive for all clinical psychologists who take their obligations seriously.—*E. J. S.*

Lewis, Hilda. *Deprived children.* New York: Oxford Univer. Press, 1954. Pp. xvii + 163. $1.55.

This is a report of 500 children who entered the Mersham Reception Centre in Kent, England, between October, 1947 and July, 1950. The Mersham Reception Centre was organized to serve two purposes: first, to plan what would be best for the children entering, and second, to gather data that might throw light on factors that were causing the social maladjustment of these children. The book tells not only the condition of the children when they entered the Centre, with their backgrounds, but also gives the results of a follow-up study made two years later. The follow-up study is especially interesting, both as to the methods used in obtaining data and the results.—*B. M. L.*

Lowenfeld, Margaret. *The Lowenfeld Mosaic Test.* London: Newman Neame; and New York: Psychological Corp., 1954. Pp. 360 + 144 color plates (boxed). $10.00.

Lowenfeld's manual for her well-known Mosaic Test deals comprehensively with its nature and administration, the definition and classification of the designs produced, and the clinical implications drawn from its results. The standard directions for adminis-

Note.—The reviews were prepared by the Editor and Advisory Editors, who may be identified by their initials.

tration, and especially the definitions and illustrations of the terms used to describe the designs, are useful to both clinical and research workers. At last they now will have a uniform method for communicating about the Mosaic. The 144 color plates make concepts clear that are hard to describe in words or even in uncolored illustrations. In its clinical sections, the book displays honest confusion. Numerous case studies give, by implication more than direct assertion, the "typical" designs produced by patients with many syndromes, children of various developmental stages, and people of several cultures. In discussing generalities, the author sometimes makes glowing claims—". . . the L. M. T. often makes it possible to make a rapid and certain diagnosis in cases of doubt . . ." (p. 246)—and sometimes shows awareness of the limitations of her method—". . . differential diagnosis between a severe degree of neurosis, a low level of inherent intelligence, and severe schizophrenia, is one of extreme difficulty, since in all three cases subjects are apt to . . ." (p. 146). A striking omission, all too familiar in books on new clinical techniques, is even the simplest evidence on validity, other than anecdotes. Is it too much to ask that authors at least count and report the percentages of hits and misses made by the test when predicting certain verifiable diagnoses or other events?—*L. F. S.*

McNemar, Quinn. *Psychological statistics.* (2nd Ed.) New York: Wiley, 1955. Pp. vii + 408. $6.00.

Like the first edition (*J. consult. Psychol.*, 1949, 13, 308), the revision of McNemar's *Statistics* remains a clear and precise resource for instruction and reference. Changes include some expansions and rearrangements to make the text less difficult, a new treatment of interaction in complex analysis of variance, and a short but useful chapter on distribution-free methods.—*L. F. S.*

Meehl, Paul E. *Clinical versus statistical prediction.* Minneapolis: Univer. of Minnesota Press, 1954. Pp. x + 149. $3.00.

Meehl's essay on the clinical and the statistical, of which many of us have heard rumors for several years, proves to have been worth waiting for. It is a superb piece of thinking, at once fair-minded and stimulating, profound and readable, serious and witty. First, he clears away some cobwebs of semantics that have kept clinicians and the statistically minded from thinking together constructively. Second, he analyzes with penetrating sympathy the logic of clinical activity, in a chapter which is perhaps the book's most distinctive contribution. Third, he reviews twenty empirical studies—all that can be found today—which attempt to compare the values of clinical and actuarial prediction. It is something of a shock to find how meager is the evidence. Some of the twenty studies are manifestly imperfect in design, others deal with issues such as academic prediction that most clinicians would regard as peripheral. The studies find in favor of the actuarial method, at least by a shade, but they point clearly to the need for more and better research. Proponents of both the statistical and clinical approaches need to read this little volume. And let neither side misperceive the author as being in the camp of the enemy; he has much to teach all of us. Read and see for yourself.—*L. F. S.*

Rubinstein, Eli A., & Lorr, Maurice. (Eds.) *Survey of clinical practice in psychology.* New York: International Universities Press, 1954. Pp. xvii + 363. $6.00.

The purpose of this volume is to offer "a comprehensive description of the function and role of the clinical psychologist." It does this by getting some 32 contributors from various clinical fields "to discuss the relationship of clinical psychology to the particular setting in which he practices and to describe the functions of the clinician which characterize his operation in that setting or organization." All the typical clinical settings are represented, and the list of contributors reads like a Who's Who of clinical psychology. If we start in alphabetical order with the list of contributors, we find Bobbitt writing on the Public Health Service, Brewer on the Wichita Guidance Center, and Brotemarkle on the University of Pennsylvania Psychological Clinic; and so on! Such compendia are often thrown together hurriedly of second-rate contributions. This one is an exception. It is obvious that both the authors and the editors have put their hearts into their work. The result is excellent. It is of course a fine reference book, but it also makes stimulating reading. This reviewer would label it a "must" for every clinician's bookshelf.—*W. A. H.*

Schafer, Roy. *Psychoanalytic interpretation in Rorschach testing.* New York: Grune & Stratton, 1954. Pp. xiv + 446. $8.75.

Believing that ". . . much of today's most important theoretical and clinical progress in the study of personality is being made through psychoanalytic ego psychology," and that, in the Rorschach, "Score analysis is *minimal* analysis," Schafer proceeds to demonstrate that Rorschach interpretation can be tremendously enriched by the application of the concepts of psychoanalytic psychology. Interwoven throughout the book is a simplified exposition of some of the basic psychoanalytic concepts, but the application of these concepts to Rorschach interpretation is very much deeper. There is an excellent chapter on interpersonal dynamics in the test situation—a subject too often neglected in Rorschach literature—and another worth-while chapter on the theory of the Rorschach response process, also within the framework of psychoanalytic psychology. Rorschach workers will find the discussion on content analysis, called "thematic analysis" by Schafer, rewarding. A large portion of the work is devoted to psychoanalytic defense mechanisms, studied in relation to Rorschach responses; "obsessive-compulsive defense operations" are included among them. The detailed case presentations add substance to these discussions.—*M. K.*

Test

Edwards, Allen L. *Edwards Personal Preference Schedule.* College-adult. 1 form. Untimed, (40) min. Booklet ($3.00 per 25) with template, manual, pp. 18; answer sheet and profile ($2.90 per 50); specimen set (60¢). New York: Psychological Corp., 1953, 1954.

Edwards has presented an ingenious and novel instrument for personality assessment which will surely evoke wide comment, considerable clinical use, and much research. The *Schedule* is designed in terms of 15 of Murray's manifest needs—achievement, deference, order, exhibition, autonomy, etc.—each of which is paired twice with each of the others. The 225 forced-choice items each consist of a pair of alternatives carefully equated for social desirability. As a result of this unusual and valuable matching, only two of the need scores have correlations with social desirability significantly above zero and these two are low (.32). The *PPS* thereby sidesteps that pitfall of many questionnaires, ego involvement. Because the need scales are short, the modest reliabilities are not unexpected: internal consistencies range from .60 to .87, and retest correlations from .74 to .88. The subscore intercorrelations are low. The validity of such a schedule is not easily expressed in simple terms, but the manual contains interesting data on clinical observations, and on relationships with ratings and other questionnaires. It is a long time since this reviewer has seen a questionnaire that seems to possess such potentialities for use and research.—*L. F. S.*

Books Received

Adams, Donald K. *The anatomy of personality.* Doubleday Papers in Psychology, No. 11. Garden City, N. Y.: Doubleday, 1954. Pp. vii + 44 (paper). 85¢.

Allen, Robert M. *Elements of Rorschach interpretation.* New York: International Universities Press, 1954. Pp. 242. $4.00.

Balint, Alice. *The early years of life.* New York: Basic Books, 1954. Pp. viii + 149. $3.00.

Bernstein, Arnold. *On the nature of psychotherapy.* Doubleday Papers in Psychology, No. 12. Garden City, N. Y.: Doubleday, 1954. Pp. ix + 36 (paper). 85¢.

Blum, Lucille Hollander, Davidson, Helen H., & Fieldsteel, Nina D. *A Rorschach workbook.* New York: International Universities Press, 1954. Pp. iv + 167 (paper). $2.00.

Coleman, Lester L. *Freedom from fear.* New York: Hawthorn, 1954. Pp. 285. $3.95.

Cowell, Roberta. *Roberta Cowell's story.* New York: British Book Centre, 1954. Pp. ix + 208. $3.00.

Deutsch, Felix, & Murphy, William F. *The clinical interview.* Vol. 1. *Diagnosis.* New York: International Universities Press, 1954. Pp. 613. $10.00.

Duvall, Evelyn Millis. *In-laws: pro & con.* New York: Association Press, 1954. Pp. viii + 400. $3.95.

Ferm, Vergilius. *A dictionary of pastoral psychology.* New York: Philosophical Library, 1954. Pp. xi + 336. $6.00.

Gondor, Emery I. *Art and play therapy.* Doubleday Papers in Psychology, No. 10. Garden City, N. Y.: Doubleday, 1954. Pp. x + 61 (paper). 85¢.

Hadamard, Jacques. *The psychology of invention in the mathematical field.* New York: Dover Publications, 1954. Pp. xiii + 145. $1.25 paper; $2.50 cloth.

Hall, Calvin S. *A primer of Freudian psychology.* Cleveland and New York: World Publishing Co., 1954. Pp. xii + 137. $2.50.

Heckler, Edwin L. *The meat packing industry.* Vocational and Professional Monographs, No. 46. Cambridge, Mass.: Bellman Publishing Co., 1954. Pp. 23 (paper). $1.00.

Husquinet, Albert. *L'adaptation scolaire et familiale des jeunes garçons de 12 à 14 ans.* Paris: Société de l'Edition "Les Belles Lettres," 1954. Pp. 199 + 17.

Keller, Fred S. *Learning: reinforcement theory.* Doubleday Papers in Psychology, No. 13. Garden City, N. Y.: Doubleday, 1954. Pp. ix + 37 (paper). 85¢.

Kerman, Edward F. *What is electroshock therapy?* New York: Exposition Press, 1954. Pp. 152. $3.50.

Macfarlane, Jean W., Allen, Lucile, & Honzik, Marjorie P. *A developmental study of the behavior problems of normal children between twenty-one months and fourteen years.* Berkeley and Los Angeles: Univer. of California Press, 1954. Pp. vii + 222 (paper). $2.25.

Nottingham, Elizabeth K. *Religion and society.* Doubleday Short Studies in Psychology, No. 5. Garden City, N. Y.: Doubleday, 1954. Pp. x + 84. 95¢.

Robinson, Mary Frances, & Freeman, Walter. *Psychosurgery and the self.* New York: Grune & Stratton, 1954. Pp. ix + 118. $3.00.

Sasha, Adrian Waldo. *The laws of life.* San Francisco: Living Knowledge Foundation, 1954. Pp. 224. $5.00.

Scott, Edward M. *Harvest time.* New York: Pageant Press, 1954. Pp. 137. $3.00.

Stalnaker, John M., & Dykman, Ross A. *Admission requirements, 1955.* Chicago: Association of American Medical Colleges, 1954. Pp. iii + 195. $2.00.

Ulett, George. *Rorschach introductory manual.* Beverly Hill, Calif.: Western Psychological Services, 1955. Pp. 49 (paper). $3.50.

Wolfenstein, Martha. *Children's humor.* Glencoe, Ill.: Free Press, 1954. Pp. 224. $3.75.

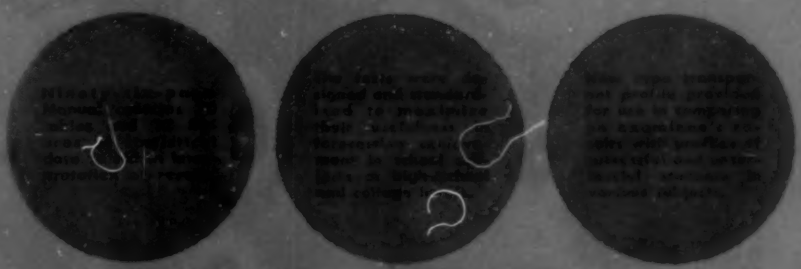
after 7 years of research

multiple aptitude tests

DEVISED BY
David Segel, Ph. D. (Stanford), Specialist in Tests and Measurements, U. S. Office of Education
Evelyn Raskin, Ph. D. (Minnesota), Assistant Professor of Psychology, Brooklyn College

DESIGNED TO
provide valid and reliable information for students in Grades 7-13 and adults in nine primary aptitude areas and four basic factors. The nine tests are: Test 1, Word Meaning; Test 2, Paragraph Meaning; Test 3, Language Usage; Test 4, Routine Clerical Facility; Test 5, Arithmetic Reasoning; Test 6, Arithmetic Computation; Test 7, Applied Science and Mechanics; Test 8, Spatial Relations–Two Dimensions; and Test 9, Spatial Relations–Three Dimensions. The four factors are: I, Verbal Comprehension; II, Perceptual Speed; III, Numerical Reasoning; and IV, Spatial Visualization.

RESULTS HELP TO
furnish information which will (1) help examinees understand their aptitudes better so as to make realistic vocational plans and choose the school subjects in which they will enjoy the greatest success, and (2) aid counselors and other school personnel in guiding individual students and adjusting curricular offerings to their need.



We invite you to inspect these tests. Order your specimen set now. Specimen sets include one copy of each of the nine tests, 96-page manual, nine scoring keys, extended profile sheet, transparent profile sheet, class record sheet, and both answer sheets.
The price postpaid is $1.75.

CALIFORNIA TEST BUREAU
Los Angeles, Calif. • New Cumberland, Penn. • Madison, Wis. • Dallas, Texas

A new Test Division publication

WAIS

Revision of Form I of the *Wechsler-Bellevue Intelligence Scale*. Product of four years of research by David Wechsler and the staff of The Psychological Corporation.

Wechsler Adult Intelligence Scale

The Tests

Same Structure. Six verbal tests and five performance tests, yielding Verbal, Performance, and Full Scale IQs.

Revised Content. Improvements in every test except Digit Span; entirely new Vocabulary test. Range of difficulty extended, especially at feebleminded level. Obsolete and less efficient items removed; new items carefully tried out before standardization testing began.

Improved Administration. Directions and scoring more precise and generally simpler, following suggestions of experienced testers. Materials and Record Form designed for efficiency. Manual provides complete directions and basic standardization data on sampling, reliability, and intercorrelations, but does not duplicate *The Measurement of Adult Intelligence* as a text.

The Standardization

Census-Controlled Sample. Seventeen hundred cases from ages 16 to 64, chosen so that each age group within each of four major geographic regions reflects 1950 census facts regarding education, occupation, and urban or rural residence. Proportional non-white population included.

Competent Supervisors and Examiners. Field testing supervised by experienced psychologists in 24 centers. People tested in over a hundred communities.

Scaled Scores and Deviation IQs. Scaled scores for each test compare an individual's performance with that of a standard reference group of 500 between the ages of 20 and 34. IQ tables computed separately for each age group. Supplementary data for older people from age 65 to over 75.

Price: Complete Set, including all material, Manual, and 25 Record Forms, $21.00 postpaid in USA. Manual alone, $2.75. Specify *WAIS* to distinguish from the *Wechsler-Bellevue Scales* and the *WISC*. (See catalog for price of carrying case.)

Status of W-B I and W-B II: Record Forms and replacement parts for Form I will be kept in stock indefinitely. Complete Sets of Form I will be available as long as sales volume supports production and inventory. *WAIS cannot be administered with Form I or II equipment.* Form II has not been revised and there are no plans for revision at this time; it will remain available indefinitely. *The Measurement of Adult Intelligence* (published by Williams and Wilkins) probably will be available as long as there is demand or until a substitute or revision appears.

Ψ **The Psychological Corporation**
522 Fifth Avenue, New York 36, N. Y.